

16869K-103400 US

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office

出願年月日

Date of Application:

2003年 1月20日

出願番号

Application Number:

特願2003-011595

[ST.10/C]:

[JP2003-011595]

出願人

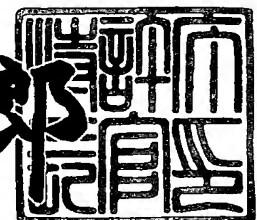
Applicant(s):

株式会社日立製作所

2003年 6月13日

特許庁長官
Commissioner,
Japan Patent Office

太田 信一郎



出証番号 出証特2003-3046499

【書類名】 特許願

【整理番号】 HI020927

【提出日】 平成15年 1月20日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 3/06

【発明者】

【住所又は居所】 神奈川県小田原市中里322番地2号 株式会社日立製作所 R A I D システム事業部内

【氏名】 小笠原 裕

【発明者】

【住所又は居所】 神奈川県小田原市中里322番地2号 株式会社日立製作所 R A I D システム事業部内

【氏名】 ▲高▼田 豊

【発明者】

【住所又は居所】 神奈川県小田原市中里322番地2号 株式会社日立製作所 R A I D システム事業部内

【氏名】 小林 直孝

【発明者】

【住所又は居所】 神奈川県横浜市戸塚区戸塚町5030番地 株式会社日立製作所 ソフトウェア事業部内

【氏名】 志川 甚一

【発明者】

【住所又は居所】 神奈川県横浜市戸塚区戸塚町5030番地 株式会社日立製作所 ソフトウェア事業部内

【氏名】 雜賀 信之

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社日立製作所

【代理人】

【識別番号】 100071283

【弁理士】

【氏名又は名称】 一色 健輔

【選任した代理人】

【識別番号】 100084906

【弁理士】

【氏名又は名称】 原島 典孝

【選任した代理人】

【識別番号】 100098523

【弁理士】

【氏名又は名称】 黒川 恵

【選任した代理人】

【識別番号】 100112748

【弁理士】

【氏名又は名称】 吉田 浩二

【選任した代理人】

【識別番号】 100110009

【弁理士】

【氏名又は名称】 青木 康

【手数料の表示】

【予納台帳番号】 011785

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 記憶デバイス制御装置にソフトウェアをインストールする方法
、記憶デバイス制御装置の制御方法、及び記憶デバイス制御装置

【特許請求の範囲】

【請求項1】 情報処理装置から送信されるファイル単位でのデータ入出力要求を第一のネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有するチャネル制御部と、

前記I/Oプロセッサから送信されるI/O要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

前記チャネル制御部および前記ディスク制御部を通信可能に接続する第二のネットワークと

を備える記憶デバイス制御装置にソフトウェアをインストールする方法であって

前記ファイルアクセス処理部を機能させるためのソフトウェアを、前記第二のネットワークを通じて前記チャネル制御部と通信することにより前記記憶デバイスに書き込むステップを備えることを特徴とする記憶デバイス制御装置にソフトウェアをインストールする方法。

【請求項2】 情報処理装置から送信されるファイル単位でのデータ入出力要求を第一のネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有するチャネル制御部と、

前記I/Oプロセッサから送信されるI/O要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

前記チャネル制御部および前記ディスク制御部を通信可能に接続する第二のネットワークと

を備える記憶デバイス制御装置にソフトウェアをインストールする方法であって

前記第二のネットワークを通じて前記チャネル制御部および前記ディスク制御

部に設けられている不揮発性メモリにファームウェアを書き込むステップを備えることを特徴とする記憶デバイス制御装置にソフトウェアをインストールする方法。

【請求項3】 請求項1に記載の記憶デバイス制御装置にソフトウェアをインストールする方法において、

前記記憶デバイスには、前記チャネル制御部の前記ファイルアクセス処理部を機能させるためのソフトウェアを記憶するための記憶領域が割り当てられていること

を特徴とする記憶デバイス制御装置にソフトウェアをインストールする方法。

【請求項4】 請求項1に記載の記憶デバイス制御装置にソフトウェアをインストールする方法において、

前記ファイルアクセス処理部を機能させるための前記ソフトウェアは、ファイルシステムの機能を提供するオペレーティングシステムの機能を実現するためのソフトウェアであること

を特徴とする記憶デバイス制御装置にソフトウェアをインストールする方法。

【請求項5】 請求項1に記載の記憶デバイス制御装置にソフトウェアをインストールする方法において、

前記第二のネットワークにはコンピュータが接続されており、

前記ソフトウェアは、前記チャネル制御部が前記コンピュータと通信することにより、前記コンピュータから前記記憶デバイスに書き込まれることを特徴とする記憶デバイス制御装置にソフトウェアをインストールする方法。

【請求項6】 請求項1に記載の記憶デバイス制御装置にソフトウェアをインストールする方法において、

前記第二のネットワークにはコンピュータが接続されており、

前記コンピュータが、前記通信を行う前記チャネル制御部を識別するための情報を記憶するステップと、

前記コンピュータが前記情報を記憶している特定の前記チャネル制御部と通信することにより、前記コンピュータから前記ソフトウェアを前記記憶デバイスに書き込むステップと

を備えることを特徴とする記憶デバイス制御装置にソフトウェアをインストールする方法。

【請求項7】 請求項2に記載の記憶デバイス制御装置にソフトウェアをインストールする方法において、

前記ファームウェアは、前記第二のネットワークに接続されるコンピュータから前記チャネル制御部および前記ディスク制御部に送信されることを特徴とする記憶デバイス制御装置にソフトウェアをインストールする方法。

【請求項8】 請求項2に記載の記憶デバイス制御装置にソフトウェアをインストールする方法において、

前記第二のネットワークにはコンピュータが接続されており、

前記コンピュータが、前記通信を行う前記チャネル制御部及び前記ディスク制御部を識別するための情報を記憶するステップと、

前記コンピュータが前記情報を記憶している特定の前記チャネル制御部及び前記ディスク制御部と通信することにより、前記コンピュータから前記ファームウェアを前記チャネル制御部及び前記ディスク制御部に書き込むステップとを備えることを特徴とする記憶デバイス制御装置にソフトウェアをインストールする方法。

【請求項9】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有する複数のチャネル制御部を含んで構成され、

前記各チャネル制御部はフェイルオーバーを行うためのグループに類別されている記憶デバイス制御装置の制御方法であって、

前記記憶デバイス制御装置が、前記グループ内の全回路基板が同一系統の電源に接続されていないかどうかを検査するステップと、

前記グループ内の全回路基板が同一系統の電源に接続されている場合には、ユーザインターフェースに警告を出力するステップとを備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項10】 情報処理装置から送信されるファイル単位でのデータ入出

力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有する複数のチャネル制御部を含んで構成され、

前記各チャネル制御部はフェイルオーバを行うためのグループに類別されている記憶デバイス制御装置であって、

前記記憶デバイス制御装置が、前記グループ内の全回路基板が同一系統の電源に接続されていないかどうかを検査する手段と、

前記グループ内の全回路基板が同一系統の電源に接続されている場合には、ユーザインターフェースに警告を出力する手段と
を備えることを特徴とする記憶デバイス制御装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、記憶デバイス制御装置にソフトウェアをインストールする方法、記憶デバイス制御装置の制御方法、及び記憶デバイス制御装置に関する。

【0002】

【従来の技術】

近年コンピュータシステムで取り扱われるデータ量が急激に増加している。このようなデータを管理するためのストレージシステムとして、最近ではミッドレンジクラスやエンタープライズクラスと呼ばれるような、巨大な記憶資源を提供するRAID (Redundant Arrays of Inexpensive Disks) 方式で管理された大規模なストレージシステムが注目されている。また、かかる膨大なデータを効率よく利用し管理するために、ディスクアレイ装置等のストレージシステムと情報処理装置とを専用のネットワーク (Storage Area Network、以下SANと記す) で接続し、ストレージシステムへの高速かつ大量なアクセスを実現する技術が開発されている。

一方、ストレージシステムと情報処理装置とをTCP/IP (Transmission Control Protocol/Internet Protocol) プロトコル等を用いたネットワークで相互に接続し、情報処理装置からのファイルレベルでのアクセスを実現するNAS

(Network Attached Storage) と呼ばれるストレージシステムが開発されている

【0003】

【特許文献1】

特開2002-351703号公報

【0004】

【発明が解決しようとする課題】

しかしながら従来のNASは、TCP/IP通信機能及びファイルシステム機能を持たないストレージシステムに、TCP/IP通信機能及びファイルシステム機能を持った情報処理装置を接続させることにより実現されていた。そのため、上記接続される情報処理装置の設置スペースが必要であった。また上記情報処理装置とストレージシステムとの間は、高速に通信を行う必要性からSANで接続されていることが多く、そのための通信制御機器や通信制御機能を備える必要もあった。またストレージシステムをNASとして機能させるためには、上記TCP/IP通信機能及びファイルシステム機能を持たないストレージシステムと、TCP/IP通信機能及びファイルシステム機能を持った情報処理装置とのそれぞれに対してソフトウェアをインストールする必要があり、さらにそれらのソフトウェアを連携させるための様々な設定を行う必要があった。

本発明は上記課題を鑑みてなされたものであり、記憶デバイス制御装置にソフトウェアをインストールする方法、記憶デバイス制御装置の制御方法、及び記憶デバイス制御装置を提供することを主たる目的とする。

【0005】

【課題を解決するための手段】

上記課題を解決するために、本発明に係る記憶デバイス制御装置にソフトウェアをインストールする方法は、情報処理装置から送信されるファイル単位でのデータ入出力要求を第一のネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有するチャネル制御部と、前記I/Oプロセッサから送信されるI/O要求に応じて前記記憶デバイスに対するデ

ータ入出力を実行するディスク制御部と、前記チャネル制御部および前記ディスク制御部を通信可能に接続する第二のネットワークとを備える記憶デバイス制御装置にソフトウェアをインストールする方法であって、前記ファイルアクセス処理部を機能させるためのソフトウェアを、前記第二のネットワークを通じて前記チャネル制御部と通信することにより前記記憶デバイスに書き込むステップを備える。

【0006】

なお、前記情報処理装置とは、前記構成の前記記憶デバイス制御装置を備えて構成されるストレージシステムにLANやSANを介してアクセスする、例えば、パーソナルコンピュータやメインフレームコンピュータである。ファイルアクセス処理部の機能はCPU上で実行されるオペレーティングシステムおよびこのオペレーティングシステム上で動作する例えばNFS (Network File System) 等のソフトウェアによって提供される。記憶デバイスは例えばハードディスク装置などのディスクドライブである。I/Oプロセッサは例えばファイルアクセス処理部のハードウェア要素である前記CPUとは独立したIC (Integrated Circuit) をハードウェア要素とし、ファイルアクセス処理部とディスク制御部との間の通信を制御する。ディスク制御部は、記憶デバイスに対してデータの書き込みや読み出しを行う。

【0007】

またファームウェアやファイルアクセス処理部を機能させるためのソフトウェアが記憶デバイスにインストールされることにより、ストレージシステムは情報処理装置に対してNASとしての機能を提供することができる。

【0008】

その他、本願が開示する課題、及びその解決方法は、発明の実施の形態の欄、及び図面により明らかにされる。

【0009】

【発明の実施の形態】

以下、本発明の実施の形態について図面を用いて詳細に説明する。

まず、本実施の形態に係るストレージシステム600の全体構成を示すプロッ

ク図を図1に示す。

【0010】

====全体構成例====

ストレージシステム600は、記憶デバイス制御装置100と記憶デバイス300とを備えている。記憶デバイス制御装置100は、情報処理装置200から受信したコマンドに従って記憶デバイス300に対する制御を行う。例えば情報処理装置200からデータの入出力要求を受信して、記憶デバイス300に記憶されているデータの入出力のための処理を行う。データは、記憶デバイス300が備えるディスクドライブにより提供される物理的な記憶領域上に論理的に設定される記憶領域である論理ボリューム（Logical Unit）（以下、LUと記す）に記憶されている。また記憶デバイス制御装置100は、情報処理装置200との間で、ストレージシステム600を管理するための各種コマンドの授受も行う。

【0011】

情報処理装置200はCPU（Central Processing Unit）やメモリを備えたコンピュータである。情報処理装置200が備えるCPUにより各種プログラムが実行されることにより様々な機能が実現される。情報処理装置200は、例えばパーソナルコンピュータやワークステーションであることもあるし、メインフレームコンピュータであることもある。

【0012】

図1において、情報処理装置1乃至3（200）は、LAN（Local Area Network）400を介して記憶デバイス制御装置100と接続されている。LAN400は、インターネットとすることもできるし、専用のネットワークとすることもできる。LAN400を介して行われる情報処理装置1乃至3（200）と記憶デバイス制御装置100との間の通信は、例えばTCP/IPプロトコルに従って行われる。情報処理装置1乃至3（200）からは、ストレージシステム600に対して、ファイル名指定によるデータアクセス要求（ファイル単位でのデータ入出力要求。以下、ファイルアクセス要求と記す）が送信される。

【0013】

LAN400にはバックアップデバイス910が接続されている。バックアッ

デバイス910は具体的にはMOやCD-R、DVD-RAMなどのディスク系デバイス、DATテープ、カセットテープ、オープンテープ、カートリッジテープなどのテープ系デバイスである。バックアップデバイス910は、LAN400を介して記憶デバイス制御装置100との間で通信を行うことにより、記憶デバイス300に記憶されているデータのバックアップデータを記憶する。またバックアップデバイス910は情報処理装置1(200)と接続されるようになることもできる。この場合は情報処理装置1(200)を介して記憶デバイス300に記憶されているデータのバックアップデータを取得するようになる。

【0014】

記憶デバイス制御装置100は、チャネル制御部1乃至4(110)を備える。記憶デバイス制御装置100は、チャネル制御部1乃至4(110)によりLAN400を介して情報処理装置1乃至3(200)やバックアップデバイス910との間で通信を行う。チャネル制御部1乃至4(110)は、情報処理装置1乃至3(200)からのファイルアクセス要求を個々に受け付ける。すなわち、チャネル制御部1乃至4(110)には、個々にLAN400上のネットワークアドレス(例えば、IPアドレス)が割り当てられていてそれが個別にNASとして振る舞い、個々のNASがあたかも独立したNASが存在するかのようにNASとしてのサービスを情報処理装置1乃至3(200)に提供することができる。以下、チャネル制御部1乃至4(110)をCHNと記す。このように1台のストレージシステム600に個別にNASとしてのサービスを提供するチャネル制御部1乃至4(110)を備えるように構成したことで、従来、独立したコンピュータで個々に運用されていたNASサーバが一台のストレージシステム600に集約される。そして、これによりストレージシステム600の統括的な管理が可能となり、各種設定・制御や障害管理、バージョン管理といった保守業務の効率化が図られる。

【0015】

なお、本実施の形態に係る記憶デバイス制御装置100のチャネル制御部1乃至4(110)は、後述するように、一体的にユニット化された回路基板上に形成されたハードウェア及びこのハードウェアにより実行されるオペレーティング

システム（以下、OSと記す）やこのOS上で動作するアプリケーションプログラムなどのソフトウェアにより実現される。このように本実施例のストレージシステム600では、従来ハードウェアの一部として実装されてきた機能がソフトウェアにより実現されている。このため、本実施例のストレージシステム600では柔軟性に富んだシステム運用が可能となり、多様で変化の激しいユーザニーズによりきめ細かなサービスを提供することが可能となる。

【0016】

情報処理装置3乃至4（200）はSAN（Storage Area Network）500を介して記憶デバイス制御装置100と接続されている。SAN500は、記憶デバイス300が提供する記憶領域におけるデータの管理単位であるブロックを単位として情報処理装置3乃至4（200）との間でデータの授受を行うためのネットワークである。SAN500を介して行われる情報処理装置3乃至4（200）と記憶デバイス制御装置100との間の通信は、一般にファイバチャネルプロトコルに従って行われる。情報処理装置3乃至4（200）からは、ストレージシステム600に対して、ファイバチャネルプロトコルに従ってブロック単位のデータアクセス要求（以下、ブロックアクセス要求と記す）が送信される。

【0017】

SAN500にはSAN対応のバックアップデバイス900が接続されている。SAN対応バックアップデバイス900は、SAN500を介して記憶デバイス制御装置100との間で通信を行うことにより、記憶デバイス300に記憶されているデータのバックアップデータを記憶する。

【0018】

記憶デバイス制御装置100は、チャネル制御部5乃至6（110）を備える。記憶デバイス制御装置100は、チャネル制御部5乃至6（110）によりSAN500を介して情報処理装置3乃至4（200）及びSAN対応バックアップデバイス900との間の通信を行う。以下、チャネル制御部5乃至6（110）をCHFと記す。

【0019】

また情報処理装置5（200）は、LAN400やSAN500等のネットワ

ークを介さずに記憶デバイス制御装置100と接続されている。情報処理装置5(200)としては例えばメインフレームコンピュータとすることができる。情報処理装置5(200)と記憶デバイス制御装置100との間の通信は、例えばFICON(Fibre Connection)(登録商標)やESCON(Enterprise System Connection)(登録商標)、ACONARC(Advanced Connection Architecture)(登録商標)、FIBARC(Fibre Connection Architecture)(登録商標)などの通信プロトコルに従って行われる。情報処理装置5(200)からは、ストレージシステム600に対して、これらの通信プロトコルに従ってブロックアクセス要求が送信される。

【0020】

記憶デバイス制御装置100は、チャネル制御部7乃至8(110)により情報処理装置5(200)との間で通信を行う。以下、チャネル制御部7乃至8(110)をCHAと記す。

【0021】

SAN500にはストレージシステム600の設置場所(プライマリサイト)とは遠隔した場所(セカンダリサイト)に設置される他のストレージシステム610が接続している。ストレージシステム610は、後述するレプリケーション又はリモートコピーの機能におけるデータの複製先の装置として利用される。なお、ストレージシステム610はSAN500以外にもATMなどの通信回線によりストレージシステム600に接続していることもある。この場合には例えばチャネル制御部110として上記通信回線を利用するためのインターフェース(チャネルエクステンダ)を備えるチャネル制御部110が採用される。

【0022】

本実施例によれば、ストレージシステム600内にCHN110、CHF110、CHA110を混在させて装着させることにより、異種ネットワークに接続されるストレージシステムを実現できる。具体的には、ストレージシステム600は、CHN110を用いてLAN140に接続し、かつCHF110を用いてSAN500に接続するという、SAN-NAS統合ストレージシステムである。

【0023】

====記憶デバイス====

記憶デバイス300は多数のディスクドライブ（物理ディスク）を備えており、情報処理装置200に対して記憶領域を提供する。データは、ディスクドライブにより提供される物理的な記憶領域上に論理的に設定される記憶領域であるLUに記憶されている。ディスクドライブとしては、例えばハードディスク装置やフレキシブルディスク装置、半導体記憶装置等様々なものを用いることができる。なお、記憶デバイス300は例えば複数のディスクドライブによりディスクアレイを構成するようにすることもできる。この場合、情報処理装置200に対して提供される記憶領域は、RAIDにより管理された複数のディスクドライブにより提供されるようにすることもできる。

【0024】

記憶デバイス制御装置100と記憶デバイス300との間は図1のように直接に接続される形態とすることもできるし、ネットワークを介して接続するようにすることもできる。さらに記憶デバイス300は記憶デバイス制御装置100と一体として構成されることもできる。

記憶デバイス300に設定されるLUには、情報処理装置200からアクセス可能なユーザLUや、チャネル制御部110の制御のために使用されるシステムLU等がある。システムLUにはCHN110で実行されるオペレーティングシステムも格納される。また各LUにはチャネル制御部110が対応付けられている。これによりチャネル制御部110毎にアクセス可能なLUが割り当てられている。また上記対応付けは、複数のチャネル制御部110で一つのLUを共有するようにすることもできる。なお以下において、ユーザLUやシステムLUをユーザディスク、システムディスク等とも記す。また、複数のチャネル制御部110で共有されるLUを共有LUあるいは共有ディスクと記す。

【0025】

====記憶デバイス制御装置====

記憶デバイス制御装置100はチャネル制御部110、共有メモリ120、キヤツシユメモリ130、ディスク制御部140、管理端末160、接続部150

を備える。

【0026】

チャネル制御部110は情報処理装置200との間で通信を行うための通信インターフェースを備え、情報処理装置200との間でデータ入出力コマンド等を授受する機能を備える。例えばCHN110は情報処理装置1乃至3(200)からのファイルアクセス要求を受け付ける。そしてファイルの記憶アドレスやデータ長等を求めて、ファイルアクセス要求に対応するI/O要求を出力することにより、記憶デバイス300へのアクセスを行う。これによりストレージシステム600はNASとしてのサービスを情報処理装置1乃至3(200)に提供することができる。なおI/O要求にはデータの先頭アドレス、データ長、読み出し又は書き込み等のアクセスの種別が含まれている。またデータの書き込みの場合にはI/O要求には書き込みデータが含まれているようにすることもできる。I/O要求の出力は、後述するI/Oプロセッサ119により行われる。またCHF110は情報処理装置3乃至4(200)からのファイバチャネルプロトコルに従ったロックアクセス要求を受け付ける。これによりストレージシステム600は高速アクセス可能なデータ記憶サービスを情報処理装置3乃至4(200)に対して提供することができる。またCHA110は情報処理装置5(200)からのFICONやESCON、ACONARC、FIBARC等のプロトコルに従ったロックアクセス要求を受け付ける。これによりストレージシステム600は情報処理装置5(200)のようなメインフレームコンピュータに対してもデータ記憶サービスを提供することができる。

【0027】

各チャネル制御部110は管理端末160と共に内部LAN151で接続されている。これによりチャネル制御部110に実行させるマイクロプログラム等を管理端末160から送信しインストールすることが可能となっている。チャネル制御部110の構成については後述する。

接続部150はチャネル制御部110、共有メモリ120、キャッシュメモリ130、ディスク制御部140を相互に接続する。チャネル制御部110、共有メモリ120、キャッシュメモリ130、ディスク制御部140間でのデータや

コマンドの授受は接続部150を介することにより行われる。接続部150は例えば高速スイッチングによりデータ伝送を行う超高速クロスバスイッチなどの高速バスで構成される。チャネル制御部110同士が高速バスで接続されていることで、個々のコンピュータ上で動作するNASサーバがLANを通じて接続する従来の構成に比べてチャネル制御部110間の通信パフォーマンスが大幅に向かっている。またこれにより高速なファイル共有機能や高速フェイルオーバなどが可能となる。

【0028】

共有メモリ120及びキャッシュメモリ130は、チャネル制御部110、ディスク制御部140により共有される記憶メモリである。共有メモリ120は主に制御情報やコマンド等を記憶するために利用されるのに対し、キャッシュメモリ130は主にデータを記憶するために利用される。

例えば、あるチャネル制御部110が情報処理装置200から受信したデータ入出力コマンドが書き込みコマンドであった場合には、当該チャネル制御部110は書き込みコマンドを共有メモリ120に書き込むと共に、情報処理装置200から受信した書き込みデータをキャッシュメモリ130に書き込む。一方、ディスク制御部140は共有メモリ120を監視しており、共有メモリ120に書き込みコマンドが書き込まれたことを検出すると、当該コマンドに従ってキャッシュメモリ130から書き込みデータを読み出して記憶デバイス300に書き込む。

【0029】

またあるチャネル制御部110が情報処理装置200から受信したデータ入出力コマンドが読み出しコマンドであった場合には、当該チャネル制御部110は読み出しコマンドを共有メモリ120に書き込むと共に、読み出し対象となるデータがキャッシュメモリ130に存在するかどうかを調べる。ここでキャッシュメモリ130に存在すれば、チャネル制御部110はそのデータを情報処理装置200に送信する。一方、読みだし対象となるデータがキャッシュメモリ130に存在しない場合には、共有メモリ120を監視することにより読み出しコマンドが共有メモリ120に書き込まれたことを検出したディスク制御部140が、

記憶デバイス300から読みだし対象となるデータを読み出してこれをキャッシュメモリ130に書き込むと共に、その旨を共有メモリ120に書き込む。そして、チャネル制御部110は共有メモリ120を監視することにより読みだし対象となるデータがキャッシュメモリ130に書き込まれたことを検出すると、そのデータを情報処理装置200に送信する。

【0030】

なお、このようにチャネル制御部110からディスク制御部140に対するデータの書き込みや読み出しの指示を共有メモリ120を介在させて間接に行う構成の他、例えばチャネル制御部110からディスク制御部140に対してデータの書き込みや読み出しの指示を共有メモリ120を介さずに直接に行う構成とすることもできる。

【0031】

ディスク制御部140は記憶デバイス300の制御を行う。例えば上述のように、チャネル制御部110が情報処理装置200から受信したデータ書き込みコマンドに従って記憶デバイス300へデータの書き込みを行う。また、チャネル制御部110により送信された論理アドレス指定によるLUへのデータアクセス要求を、物理アドレス指定による物理ディスクへのデータアクセス要求に変換する。記憶デバイス300における物理ディスクがRAIDにより管理されている場合には、RAID構成に従ったデータのアクセスを行う。またディスク制御部140は、記憶デバイス300に記憶されたデータの複製管理の制御やバックアップ制御を行う。さらにディスク制御部140は、災害発生時のデータ消失防止（ディザスタリカバリ）などを目的としてプライマリサイトのストレージシステム600のデータの複製をセカンダリサイトに設置された他のストレージシステム610にも記憶する制御（レプリケーション機能、又はリモートコピー機能）なども行う。

【0032】

各ディスク制御部140は管理端末160と共に内部LAN151で接続されており、相互に通信を行うことが可能である。これにより、ディスク制御部140に実行させるマイクロプログラム等を管理端末160から送信しインストール

することが可能となっている。ディスク制御部140の構成については後述する。

【0033】

本実施例においては、共有メモリ120及びキャッシュメモリ130がチャネル制御部110及びディスク制御部140に対して独立に設けられていることについて記載したが、本実施例はこの場合に限られるものでなく、共有メモリ120又はキャッシュメモリ130がチャネル制御部110及びディスク制御部140の各々に分散されて設けられることも好ましい。この場合、接続部150は、分散された共有メモリ又はキャッシュメモリを有するチャネル制御部110及びディスク制御部140を相互に接続させることになる。

【0034】

====管理端末====

管理端末160はストレージシステム600を保守・管理するためのコンピュータである。管理端末160を操作することにより、例えば記憶デバイス300内の物理ディスク構成の設定や、LUの設定、チャネル制御部110において実行されるマイクロプログラムのインストール等を行うことができる。ここで、記憶デバイス300内の物理ディスク構成の設定としては、例えば物理ディスクの増設や減設、RAID構成の変更（RAID1からRAID5への変更等）等を行うことができる。さらに管理端末160からは、ストレージシステム600の動作状態の確認や故障部位の特定、チャネル制御部110で実行されるオペレーティングシステムのインストール等の作業を行うこともできる。また管理端末160はLANや電話回線等で外部保守センタと接続されており、管理端末160を利用してストレージシステム600の障害監視を行ったり、障害が発生した場合に迅速に対応することも可能である。障害の発生は例えばOSやアプリケーションプログラム、ドライバソフトウェアなどから通知される。この通知はHTTPプロトコルやSNMP（Simple Network Management Protocol）、電子メールなどにより行われる。これらの設定や制御は、管理端末160で動作するWebサーバが提供するWebページをユーザインターフェースとしてオペレータなどにより行われる。オペレータ等は、管理端末160を操作して障害監視する対象や

内容の設定、障害通知先の設定などを行うこともできる。

【0035】

管理端末160は記憶デバイス制御装置100に内蔵されている形態とともにできるし、外付けされている形態とすることもできる。また管理端末160は、記憶デバイス制御装置100及び記憶デバイス300の保守・管理を専用に行うコンピュータとすることもできるし、汎用のコンピュータに保守・管理機能を持たせたものとすることもできる。

【0036】

管理端末160の構成を示すブロック図を図2に示す。

管理端末160は、CPU161、メモリ162、ポート163、記録媒体読取装置164、入力装置165、出力装置166、記憶装置168を備える。

【0037】

CPU161は管理端末160の全体の制御を司るもので、メモリ162に格納されたプログラム162cを実行することにより上記Webサーバとしての機能等を実現する。メモリ162には、物理ディスク管理テーブル162aとLU管理テーブル162bとプログラム162cとが記憶されている。

【0038】

物理ディスク管理テーブル162aは、記憶デバイス300に備えられる物理ディスク（ディスクドライブ）を管理するためのテーブルである。物理ディスク管理テーブル162aを図3に示す。図3においては、記憶デバイス300が備える多数の物理ディスクのうち、ディスク番号#001乃至#006までが示されている。それぞれの物理ディスクに対して、容量、RAID構成、使用状況が示されている。

【0039】

LU管理テーブル162bは、上記物理ディスク上に論理的に設定されるLUを管理するためのテーブルである。LU管理テーブル162bを図4に示す。図4においては、記憶デバイス300上に設定される多数のLUのうち、LU番号#1乃至#3までが示されている。それぞれのLUに対して、物理ディスク番号、容量、RAID構成が示されている。

【0040】

記録媒体読取装置164は、記録媒体167に記録されているプログラムやデータを読み取るための装置である。読み取られたプログラムやデータはメモリ162や記憶装置168に格納される。従って、例えば記録媒体167に記録されたプログラム162cを、記録媒体読取装置164を用いて上記記録媒体167から読み取って、メモリ162や記憶装置168に格納するようにすることができる。記録媒体167としてはフレキシブルディスクやCD-ROM、半導体メモリ等を用いることができる。記録媒体読取装置164は管理端末160に内蔵されている形態とすることもできるし、外付されている形態とすることもできる。記憶装置168は、例えばハードディスク装置やフレキシブルディスク装置、半導体記憶装置等である。入力装置165はオペレータ等による管理端末160へのデータ入力等のために用いられる。入力装置165としては例えばキーボードやマウス等が用いられる。出力装置166は情報を外部に出力するための装置である。出力装置166としては例えばディスプレいやプリンタ等が用いられる。ポート163は内部LAN151に接続されており、これにより管理端末160はチャネル制御部110やディスク制御部140等と通信を行うことができる。またポート163は、LAN400に接続するようにすることもできるし、電話回線に接続するようにすることもできる。

【0041】

====外観図====

次に、本実施の形態に係るストレージシステム600の外観構成を図5に示す。また、記憶デバイス制御装置100の外観構成を図6に示す。

図5に示すように、本実施の形態に係るストレージシステム600は記憶デバイス制御装置100及び記憶デバイス300がそれぞれの筐体に納められた形態をしている。記憶デバイス制御装置100の筐体の両側に記憶デバイス300の筐体が配置されている。

【0042】

記憶デバイス制御装置100は、正面中央部に管理端末160が備えられている。管理端末160はカバーで覆われており、図6に示すようにカバーを開ける

ことにより管理端末160を使用することができる。なお図6に示した管理端末160はいわゆるノート型パーソナルコンピュータの形態をしているが、どのような形態とすることも可能である。

【0043】

管理端末160の下部には、チャネル制御部110を装着するためのスロットが設けられている。各スロットにはチャネル制御部110のボードが装着される。本実施の形態に係るストレージシステム600においては、スロットは8つあり、図5及び図6には8つのスロットにチャネル制御部110が装着された状態が示されている。各スロットにはチャネル制御部110を装着するためのガイドレールが設けられている。ガイドレールに沿ってチャネル制御部110をスロットに挿入することにより、チャネル制御部110を記憶デバイス制御装置100に装着することができる。また各スロットに装着されたチャネル制御部110は、ガイドレールに沿って手前方向に引き抜くことにより取り外すことができる。また各スロットの奥手方向正面部には、各チャネル制御部110を記憶デバイス制御装置100と電気的に接続するためのコネクタが設けられている。チャネル制御部110には、CHN、CHF、CHAがあるが、いずれのチャネル制御部110もサイズやコネクタの位置、コネクタのピン配列等に互換性をもたせていているため、8つのスロットにはいずれのチャネル制御部110も装着することができる。従って、例えば8つのスロット全てにCHN110を装着するようになることもできる。また図1に示したように、4枚のCHN110と、2枚のCHF110と、2枚のCHA110とを装着するようにすることもできる。チャネル制御部110を装着しないスロットを設けるようにすることもできる。

【0044】

各スロットに装着されたチャネル制御部110は、同種の複数のチャネル制御部110でクラスタを構成する。例えば2枚のCHN110をペアとしてクラスタを構成することができる。クラスタを構成することにより、クラスタ内のあるチャネル制御部110に障害が発生した場合でも、障害が発生したチャネル制御部110がそれまで行っていた処理をクラスタ内の他のチャネル制御部110に引き継ぐようにすることができる（フェイルオーバ制御）。2枚のCHN110

でクラスタを構成している様子を示す図を図12に示すが、詳細は後述する。

【0045】

なお、記憶デバイス制御装置100は信頼性向上のため電源供給が2系統化されており、チャネル制御部110が装着される上記8つのスロットは電源系統毎に4つずつに分けられている。そこでクラスタを構成する場合には、両方の電源系統のチャネル制御部110を含むようにする。これにより、片方の電源系統に障害が発生し電力の供給が停止しても、同一クラスタを構成する他方の電源系統に属するチャネル制御部110への電源供給は継続されるため、当該チャネル制御部110に処理を引き継ぐ（フェイルオーバ）ことができる。

【0046】

なお、上述したように、チャネル制御部110は上記各スロットに装着可能なボード、すなわち同一のユニットに形成された一つのユニットとして提供されるが、上記同一のユニットは複数枚数の基板から構成されているようにすることもできる。つまり、複数枚数の基板から構成されていても、各基板が相互に接続されて一つのユニットとして構成され、記憶デバイス制御装置100のスロットに対して一体的に装着できる場合は、同一の回路基板の概念に含まれる。

【0047】

ディスク制御部140や共有メモリ120等の、記憶デバイス制御装置100を構成する他の装置については図5及び図6には示されていないが、記憶デバイス制御装置100の背面側等に装着されている。

また記憶デバイス制御装置100には、チャネル制御部110等から発生する熱を放出するためのファン170が設けられている。ファン170は記憶デバイス制御装置100の上面部に設けられる他、チャネル制御部110用スロットの上部にも設けられている。

【0048】

ところで、筐体に収容されて構成される記憶デバイス制御装置100および記憶デバイス300としては、例えばS A N対応として製品化されている従来構成の装置を利用することができる。特に上記のようにC H Nのコネクタ形状を従来構成の筐体に設けられているスロットにそのまま装着できる形状とすることで従

来構成の装置をより簡単に利用することができる。つまり本実施例のストレージシステム600は、既存の製品を利用することで容易に構築することができる。

【0049】

====チャネル制御部====

本実施の形態に係るストレージシステム600は、上述の通りCHN110により情報処理装置1乃至3(200)からのファイルアクセス要求を受け付け、NASとしてのサービスを情報処理装置1乃至3(200)に提供する。

【0050】

CHN110のハードウェア構成を図7に示す。この図に示すようにCHN110のハードウェアは一つのユニットで構成される。以下、このユニットのことをNASボードと記す。NASボードは一枚もしくは複数枚の回路基板を含んで構成される。より具体的には、NASボードは、ネットワークインターフェース部111、CPU112、メモリ113、入出力制御部114、I/O(Input/Output)プロセッサ119、NVRAM(Non Volatile RAM)115、ボード接続用コネクタ116、通信コネクタ117を備え、これらが同一のユニットに形成されて構成されている。

【0051】

ネットワークインターフェース部111は、情報処理装置200との間で通信を行うための通信インターフェースを備えている。CHN110の場合は、例えばTCP/IPプロトコルに従って情報処理装置200から送信されたファイルアクセス要求を受信する。通信コネクタ117は情報処理装置200と通信を行うためのコネクタである。CHN110の場合はLAN400に接続可能なコネクタであり、例えばイーサネット(登録商標)に対応している。

CPU112は、CHN110をNASボードとして機能させるための制御を司る。

【0052】

メモリ113には様々なプログラムやデータが記憶される。例えば図9に示すメタデータ730やロックテーブル720、また図11に示すNASマネージャ706等の各種プログラムが記憶される。メタデータ730はファイルシステム

が管理しているファイルに対応させて生成される情報である。メタデータ730には例えばファイルのデータが記憶されているLU上のアドレスやデータサイズなど、ファイルの保管場所を特定するための情報が含まれる。メタデータ730にはファイルの容量、所有者、更新時刻等の情報が含まれることもある。また、メタデータ730はファイルだけでなくディレクトリに対応させて生成されることもある。メタデータ730の例を図13に示す。メタデータ730は記憶デバイス300上の各LUにも記憶されている。

【0053】

ロックテーブル720は、情報処理装置1乃至3(200)からのファイルアクセスに対して排他制御を行うためのテーブルである。排他制御を行うことにより情報処理装置1乃至3(200)でファイルを共用することができる。ロックテーブル720を図14に示す。図14に示すようにロックテーブル720にはファイルロックテーブル721とLUロックテーブル722とがある。ファイルロックテーブル721は、ファイル毎にロックが掛けられているか否かを示すためのテーブルである。いずれかの情報処理装置200によりあるファイルがオープンされている場合に当該ファイルにロックが掛けられる。ロックが掛けられたファイルに対する他の情報処理装置200によるアクセスは禁止される。LUロックテーブル722は、LU毎にロックが掛けられているか否かを示すためのテーブルである。いずれかの情報処理装置200により、あるLUに対するアクセスが行われている場合に当該LUにロックが掛けられる。ロックが掛けられたLUに対する他の情報処理装置200によるアクセスは禁止される。

【0054】

入出力制御部114は、ディスク制御部140やキャッシュメモリ130、共有メモリ120、管理端末160との間でデータやコマンドの授受を行う。入出力制御部114はI/Oプロセッサ119やNVRAM115を備えている。I/Oプロセッサ119は例えば1チップのマイコンで構成される。I/Oプロセッサ119は上記データやコマンドの授受を制御し、CPU112とディスク制御部140との間の通信を中継する。NVRAM115はI/Oプロセッサ119の制御を司るプログラムを格納する不揮発性メモリである。NVRAM115

に記憶されるプログラムの内容は、管理端末160や、後述するNASマネージャ706からの指示により書き込みや書き換えを行うことができる。

【0055】

次にCHF110及びCHA110の構成を示す図を図8に示す。CHF110及びCHA110も、CHN110と同様に一つのユニットとして形成されている。CHN110と同様、上記ユニットは複数枚数の回路基板から構成されているようにすることもできる。またCHF110及びCHA110は、CHN110とサイズやボード接続用コネクタ116の位置、ボード接続用コネクタ116のピン配列等に互換性をもたせている。

CHF110及びCHA110は、ネットワークインターフェース部111、メモリ113、入出力制御部114、I/Oプロセッサ119、NVRAM115、ボード接続用コネクタ116、通信コネクタ117を備える。

【0056】

ネットワークインターフェース部111は、情報処理装置200との間で通信を行うための通信インターフェースを備えている。CHF110の場合は、例えばファイバチャネルプロトコルに従って情報処理装置200から送信されたブロックアクセス要求を受信する。CHA110の場合は、例えばFICON（登録商標）やESCON（登録商標）、ACONARC（登録商標）、FIBARC（登録商標）のプロトコルに従って情報処理装置200から送信されたブロックアクセス要求を受信する。通信コネクタ117は情報処理装置200と通信を行うためのコネクタである。CHF110の場合はSAN500に接続可能なコネクタであり、例えばファイバチャネルに対応している。CHA110の場合は情報処理装置5と接続可能なコネクタであり、FICON（登録商標）やESCON（登録商標）、ACONARC（登録商標）、FIBARC（登録商標）に対応している。

【0057】

入出力制御部114は、それぞれCHF110、CHA110の全体の制御を司ると共に、ディスク制御部140やキャッシュメモリ130、共有メモリ120、管理端末160との間でデータやコマンドの授受を行う。メモリ113に格

納された各種プログラムを実行することにより本実施の形態に係るCHF110及びCHA110の機能が実現される。入出力制御部114はI/Oプロセッサ119やNVRAM115を備えている。I/Oプロセッサ119は上記データやコマンドの授受を制御する。NVRAM115はI/Oプロセッサ119の制御を司るプログラムを格納する不揮発性メモリである。NVRAM115に記憶されるプログラムの内容は、管理端末160や、後述するNASマネージャ706からの指示により書き込みや書き換えを行うことができる。

【0058】

次にディスク制御部140の構成を示す図を図10に示す。

ディスク制御部140は、インターフェース部141、メモリ143、CPU142、NVRAM144、ボード接続用コネクタ145を備え、これらが一体的なユニットとして形成されている。

インターフェース部141は、接続部150を介してチャネル制御部110等との間で通信を行うための通信インターフェースや、記憶デバイス300との間で通信を行うための通信インターフェースを備えている。

CPU142は、ディスク制御部140全体の制御を司ると共に、チャネル制御部110や記憶デバイス300、管理端末160との間の通信を行う。メモリ143やNVRAM144に格納された各種プログラムを実行することにより本実施の形態に係るディスク制御部140の機能が実現される。ディスク制御部140により実現される機能としては、記憶デバイス300の制御やRAID制御、記憶デバイス300に記憶されたデータの複製管理やバックアップ制御、リモートコピー制御等である。

NVRAM144はCPU142の制御を司るプログラムを格納する不揮発性メモリである。NVRAM144に記憶されるプログラムの内容は、管理端末160や、NASマネージャ706からの指示により書き込みや書き換えを行うことができる。

またディスク制御部140はボード接続用コネクタ145を備えている。ボード接続用コネクタ145が記憶デバイス制御装置100側のコネクタと嵌合することにより、ディスク制御部140は記憶デバイス制御装置100と電気的に接

続される。

【0059】

次に、本実施の形態に係るストレージシステム600におけるソフトウェア構成図を図11に示す。

オペレーティングシステム701上では、RAIDマネージャ708、ボリュームマネージャ707、SVPマネージャ709、ファイルシステムプログラム703、ネットワーク制御部702、バックアップ管理プログラム710、障害管理プログラム705、NASマネージャ706などのソフトウェアが動作する。

【0060】

オペレーティングシステム701上で動作するRAIDマネージャ708は、RAID制御部740に対するパラメータの設定やRAID制御部740を制御する機能を提供する。RAIDマネージャ708はオペレーティングシステム701やオペレーティングシステム701上で動作する他のアプリケーション、もしくはSVPからパラメータや制御指示情報を受け付けて、受け付けたパラメータのRAID制御部740への設定や、RAID制御部指示情報に対応する制御コマンドの送信を行う。

【0061】

ここで設定されるパラメータとしては、例えば、RAIDグループを構成する記憶デバイス（物理ディスク）を定義（RAIDグループの構成情報、ストライプサイズの指定など）するためのパラメータ、RAIDレベル（例えば0, 1, 5）を設定するためのパラメータなどがある。また、RAIDマネージャ708がRAID制御部740に送信する制御コマンドとしてはRAIDの構成・削除・容量変更を指示するコマンド、各RAIDグループの構成情報を要求するコマンドなどがある。

【0062】

ボリュームマネージャ707は、RAID制御部740によって提供されるLUをさらに仮想化した仮想化論理ボリュームをファイルシステムプログラム703に提供する。1つの仮想化論理ボリュームは1以上の論理ボリュームによって

構成される。

【0063】

ファイルシステムプログラム703の主な機能は、ネットワーク制御部702が受信したファイルアクセス要求に指定されているファイル名とそのファイル名が格納されている仮想化論理ボリューム上のアドレスとの対応づけを管理することである。例えば、ファイルシステムプログラム703はファイルアクセス要求に指定されているファイル名に対応する仮想化論理ボリューム上のアドレスを特定する。

【0064】

ネットワーク制御部702は、NFS (Network File System) 711とSamba 712の2つのファイルシステムプロトコルを含んで構成される。NFS 711は、NFS 711が動作するUNIX (登録商標) 系の情報処理装置200からのファイルアクセス要求を受け付ける。一方、Samba 712はCIFS (Common Interface File System) 713が動作するWindows (登録商標) 系の情報処理装置200からのファイルアクセス要求を受け付ける。

【0065】

NASマネージャ706は、ストレージシステム600について、その動作状態の確認、設定や制御などを行うためのプログラムである。NASマネージャ706はWebサーバとしての機能も有し、情報処理装置200からストレージシステム600の設定や制御を行うための設定Webページを情報処理装置200に提供する。NASマネージャ706は、情報処理装置1乃至3(200)からのHTTP (HyperText Transport Protocol) リクエストに応じて、設定Webページのデータを情報処理装置1乃至3(200)に送信する。情報処理装置1乃至3(200)に表示された設定Webページを利用してシステムアドミニストレータなどによりストレージシステム600の設定や制御の指示が行われる。設定Webページを利用して行うことができる内容としては、例えば、LUの管理や設定(容量管理や容量拡張・縮小、ユーザ割り当て等)、上述の複製管理やリモートコピー(レプリケーション)等の機能に関する設定や制御(複製元のLUと複製先のLUの設定など)、後述のバックアップ管理プログラム710につ

いての設定や制御、冗長構成されたCHNやCHF、CHA間でのクラスタの管理（フェイルオーバさせる相手の対応関係の設定、フェイルオーバ方法など）、OSやOS上で動作するアプリケーションプログラムのバージョン管理、ウイルス検知プログラムやウイルス駆除などのデータの安全性に関する機能を提供するセキュリティ管理プログラム716の動作状態の管理や設定および更新管理（バージョン管理）などがある。NASマネージャ706は、設定Webページに対する操作に起因して情報処理装置200から送信される設定や制御に関するデータを受信してそのデータに対応する設定や制御を実行する。これにより、情報処理装置1乃至3（200）からストレージシステム600の様々な設定や制御を行うことができる。

【0066】

バックアップ管理プログラム710は、記憶デバイス300に記憶されているデータをLAN経由またはSAN経由でバックアップするためのプログラムである。バックアップ管理プログラム710はNDMP（Network Data Management Protocol）の機能を提供し、情報処理装置200で動作するNDMP対応のバックアップソフトウェアとLAN400を通じてNDMPに従った通信を行う。バックアップデバイス910が情報処理装置200にSCSI経由などで接続されている場合、バックアップされるデータは情報処理装置200に一旦取り込まれてからバックアップデバイス910に送られる。バックアップデバイス910がLANに200接続されている場合には、バックアップされるデータを、情報処理装置200を経由せずにストレージシステム600から直接バックアップデバイス910に転送することもできる。

【0067】

障害管理プログラム705は、クラスタを構成するチャネル制御部110間でのフェイルオーバ制御を行うためのプログラムである。

SVPマネージャ709は、管理端末160からの要求に応じて各種のサービスを管理端末160に提供する。例えば、LUの設定内容やRAIDの設定内容等のストレージシステム600に関する各種設定内容の管理端末160への提供や、管理端末160から入力されたストレージシステム600に関する各種設定

の反映等を行う。

セキュリティ管理プログラム716は、コンピュータウイルスの検知、侵入監視、コンピュータウイルス検知プログラムの更新管理、感染したコンピュータウイルスの駆除、ファイアウォール機能などを実現する。

【0068】

次に、2枚のCHN110でクラスタ180が構成されている様子を示す図を図12に示す。図12では、CHN1（チャネル制御部1）110とCHN2（チャネル制御部2）110とでクラスタ180が構成されている場合を示す。

上述したように、フェイルオーバ処理はクラスタ180を構成するチャネル制御部110間で行われる。つまり、例えばCHN1（110）に何らかの障害が発生し処理を継続することができなくなった場合には、CHN1（110）がそれまで行っていた処理はCHN2（110）に引き継がれる。フェイルオーバ処理は、CHN1（110）とCHN2（110）により実行される障害管理プログラム705により実行される。

【0069】

CHN1（110）及びCHN2（110）は共に障害管理プログラムを実行し、共有メモリ120に対して自己の処理が正常に行われていることを書き込むようとする。そして、相手側の上記書き込みの有無を相互に確認するようとする。相手側による書き込みが検出できない場合には、相手側に何らかの障害が発生したと判断し、フェイルオーバ処理を実行する。フェイルオーバ実行時の処理の引き継ぎは、共有LU310を介して行われる。

【0070】

また、クラスタ180を構成する各CHN110のファイルアクセス処理部は、アクセス可能な情報処理装置1乃至3（200）を管理することができる。これにより、アクセス可能な情報処理装置1乃至3（200）からファイルアクセス要求が送信されてきた場合にのみ、そのファイルアクセス要求を受け付けるようにすることができる。アクセス可能な情報処理装置1乃至3（200）の管理は、各CHN110のメモリ113に予めアクセスを許可する情報処理装置1乃至3（200）のIPアドレスのドメイン名を記憶させることにより行う。

【0071】

これにより、情報処理装置1乃至3（200）が共通のLAN400を介してストレージシステム600に接続されても、情報処理装置1乃至3（200）のそれぞれに対して排他的にLUを割り当てることができる。例えば情報処理装置1乃至3（200）がそれぞれ別の企業のコンピュータである場合に、情報処理装置1乃至3（200）のそれぞれに対して相互にデータの秘匿性を保った記憶サービスを提供することができる。

【0072】

各CHN110に対する上記設定は、管理端末160や情報処理装置1乃至3（200）から行うことができる。情報処理装置1乃至3（200）から行う場合には、CHN110上で動作するNASマネージャ706により情報処理装置1乃至3（200）に表示される設定Webページを利用して行う。

【0073】

====インストール処理=====

次に、本実施の形態に係るストレージシステム600をNASとして機能させるために必要なソフトウェアのインストールについて説明する。

ストレージシステム600をNASとして機能させるためには、CHN110で実行されるOS701をインストールする必要がある。また、CHN110やディスク制御部140で実行させるためのマイクロプログラム（ファームウェア）もインストールする必要がある。その他CHN110には、ボリュームマネージャ707、ファイルシステムプログラム703、NASマネージャ706等のアプリケーションプログラムが必要に応じてインストールされる。またディスク制御部140には、リモートコピー制御プログラム750、複製管理プログラム760等が必要に応じてインストールされる。

【0074】

OS701やアプリケーションプログラムは記憶デバイス300に設定されるシステムLUに記憶される。システムLUにはOSインストール領域や、障害検知用の記憶領域、クラスタ情報用の記憶領域などが割り当てられることがある。障害検知用の記憶領域には、OS701やアプリケーションプログラムが出力す

るダンプリスト（OS701のカーネルの異常終了や、デーモンの異常終了、複数のプロセス間で処理がループすることによる異常などに起因して出力されるコア（core）ダンプ、メモリダンプ、ディスクダンプ）などの障害管理に関する情報が格納される。クラスタ情報用の記憶領域にはCHN110間でクラスタを設定するために必要となる情報が格納される。このように、OS701やアプリケーションプログラムを格納するための記憶領域を記憶デバイス300に設けるようにより、CHN110にかかる記憶領域を設ける必要がなくなる。

【0075】

また上記障害検知用の記憶領域やクラスタ情報用の記憶領域は、それぞれ障害管理用LUや、クラスタLUとして、システムLUとは別に設けるようにすることもできる。なお、記憶デバイス300がRAID5の方式で運用されている場合には、システムLUや障害管理用LU、クラスタLUなどは1つのパリティグループに集中させるのではなく、複数のパリティグループに分散させるようにすることが好ましい。これらのLUには記憶デバイス制御装置100の運用上、重要なデータが格納されるためである。

【0076】

次に、ストレージシステム600をNASとして機能させるために必要な、OS701やマイクロプログラムのインストール手順について説明する。これらのプログラムのインストールは管理端末（コンピュータ）160から行われる。

インストールの手順を説明するためのブロック図を図15に示す。またインストールを行うために管理端末160の出力装置166に表示される画面例を図16に示す。

【0077】

図15に示すブロック図において、記憶デバイス制御装置100はLAN（第一のネットワーク）400を介して情報処理装置200と接続されており、情報処理装置200からファイルアクセス要求を受け付ける。また記憶デバイス制御装置100は管理端末160を備えている。管理端末160は、内部LAN（第二のネットワーク）151を介してCHN110及びディスク制御部140と接続されている。管理端末160には、マイクロプログラム770、ローダ771

、インストーラ772、及びOS773が記憶されている。これらのプログラムは管理端末160のメモリ162や記憶装置168に記憶されている。マイクロプログラム770には、CHN110のNVRAM115に書き込まれるものと、ディスク制御部140のNVRAM144に書き込まれるものとがある。前者はCHN110のI/Oプロセッサ119の制御を司るためのプログラムである。後者はディスク制御部140のCPU142の制御を司るためのプログラムである。ローダ771及びインストーラ772は、管理端末160に記憶されているOS773をCHN110が読み込む際に用いられるプログラムである。OS773はCHN110毎に記憶デバイス300に設けられるシステムLUにインストールされる。これらのプログラムは、管理端末160が備える記録媒体読取装置164を用いてCD-ROM等の記録媒体167から読み込むようにすることもできるし、ポート163を介して例えばインターネットからダウンロードすることにより読み込むようにすることもできる。

【0078】

また、図16に示す管理端末160の出力装置166に表示される画面例においては、スロットNo.1乃至4に新規に装着されたCHN110の設定を行う場合の例を示す。なお図16においては、スロットNo.5乃至8についても設定内容の表示がなされているが、既に設定済みのスロットに対しては、非表示にする等、オペレータからの設定入力を行えないようにすることもできる。

【0079】

図16に示すインストール設定画面には、スロットNo.欄、チャネルアダプタ設定欄、システムLU No.指定欄、ブート指定欄がある。このうちオペレータが入力できるのは、チャネルアダプタ設定欄、システムLU No.指定欄、ブート指定欄である。

【0080】

チャネルアダプタ設定欄には、記憶デバイス制御装置100のスロットに装着するチャネル制御部110の種類を入力する。チャネル制御部110の種類には、CHA、CHF、CHNがある。これらの入力は、図16の下向き三角マークで示される部分をマウスでクリックすることにより選択できるようになっている

【0081】

システムLU No.指定欄にはシステムLUの番号を入力する。システムLUとしては、記憶デバイス300に設定されているLUの中から任意のLUを選択することが可能である。システムLUはCHN110毎に指定する。なお、システムLUとして指定されたLUの記憶容量が所定の容量に満たない場合には、当該LUにOS773等のインストールを行うことができない。そのため、管理端末160は、システムLU No.指定欄に入力されたLUの容量をチェックする機能を備えている。なお、システムLUは複数のCHN110で共有するようにすることもできる。

【0082】

ブート指定欄には、チャネル制御部110のブート方法を入力する。ネットワーク起動の場合は、内部LAN151で接続された管理端末160からブートが行われる。ディスク起動の場合は、記憶デバイス300上のシステムSHIからブートが行われる。これらの指定は、図16の下向き三角マークで示される部分をマウスでクリックすることにより選択できるようになっている。

【0083】

オペレータは、スロットNo.1乃至4に対して図16に示すような入力を行つたら、OKボタンをクリックする。そうすると、管理端末160はスロットNo.1乃至4の装着されたCHN110に対して順にインストールを開始する。

【0084】

インストールの手順を示すフローチャートを図18に示す。管理端末160を操作するオペレータの指示により、管理端末160において実行されるマイクロプログラム書き換えプログラムにより、まずマイクロプログラム770の書き込み先となるCHN110とディスク制御部140のMAC (Media Access Control) アドレス (チャネル制御部110及びディスク制御部140を識別するための情報) が管理端末160の例えばメモリ162に記憶される。そして、管理端末160がMACアドレスを記憶している特定のCHN110のNVRAM115とディスク制御部140のNVRAM144にそれぞれマイクロプログラム7

70が書き込まれる(S1000)。この書き込みは、管理端末160から内部LAN151を通じて行われる。マイクロプログラム770の書き込み先は、上記インストール設定画面からの入力情報に基づいて決めるようになることができる。またマイクロプログラム770の書き込み先のMACアドレスの取得は、内部LAN151に接続されているCHN110やディスク制御部140に対してMACアドレス問い合わせコマンドを送信することにより行うことができる。なお、管理端末160により記憶されるCHN110やディスク制御部140を識別するための情報はMACアドレスに限られない。例えばIPアドレスとしてもできるし、CHN110やディスク制御部140の製造番号とすることもできる。また書き込まれるマイクロプログラム770は、CD-ROMなどの記録媒体167により提供されることもあるし、インターネット経由でダウンロードされることもある。

【0085】

このように、管理端末160にMACアドレスが記憶されるCHN110やディスク制御部140に対してマイクロプログラム770の書き込みが行われることにより、特定のCHN110やディスク制御部140に対してのみマイクロプログラム770を書き込むようにすることができる。これにより、マイクロプログラム770を書き換える必要のあるCHN110やディスク制御部140に対してのみマイクロプログラム770を書き込むことができる。また、CHN110用のマイクロプログラム770を誤ってCHA110やCHF110に書き込んでしまうことを防止することができる。

続いて、CHN110およびディスク制御部140がリセットされる。これによりインストールしたマイクロプログラム770の動作が開始する。ここでのリセットは、例えば管理端末160により行われる。リセット後、ディスク制御部140が動作を開始すると管理端末160は記憶デバイス300のLUを認識できるようになる。

【0086】

次に、管理端末160にMACアドレスが記憶されているCHN110に対して、管理端末160からローダ771の読み込み指示が送信される。ローダ77

1は、CHN110が管理端末160からインストーラ772を読み込むために実行するプログラムである。管理端末160から送信されるローダ771の読み込み指示は、CHN110のCPU112のBIOS (Basic Input/Output System)に取り込まれる。これにより管理端末160からローダ771がCHN110に読み込まれる (S1001)。

続いてローダ771は、管理端末160からインストーラ772を読み込む (S1002)。インストーラ772は内部LAN151の通信に関する機能を提供する他、ファイルシステムベースでLUをフォーマットする機能や、LUへのファイルの書き込み機能を提供する。

【0087】

次にインストーラ772は、OS773のインストール領域を確保するため、システムLUにパーティションを設定し、ファイルシステムとしてフォーマットを行う (S1003)。そして管理端末160からOS773を読み込み、上記フォーマットしたインストール領域にファイル単位で書き込む (S1004)。またインストーラ772は書き込んだOS773の設定ファイルに内部LAN151のネットワーク設定を反映させる。

さらにインストーラ772はOS773を書き込んだシステムLUにMBR (Master Boot Record) の起動コードを書き込んで、MBRを有効にする (S1005)。これによりCHN110のリセット時に自動的にOS773が起動されるようになる。

【0088】

以上の処理が完了すると、CHN110のI/Oプロセッサ119はインストールが完了した旨を、内部LAN151経由で管理端末160に通知する (S1006)。

管理端末160は、上記のインストール処理をスロットNo.1乃至4の各CHN110に対して繰り返し行う (S1007)。スロットNo.1乃至4の全てのCHN110に対して上記処理が完了したら、インストール処理が完了する。

このようなインストール方法により、ファームウェアやOS773を一連の手順により円滑に行うことが可能となる。

【0089】

続いてCHN110に対してNASとしての設定を行う。CHN110に対して行うNASの設定は、IPアドレスの付与、ユーザ領域の設定、OSの設定、クラスタの設定である。これらの設定はNASマネージャ110により行う。IPアドレスの付与は各CHN110が有する2つの通信ポートのそれぞれに対して行う。ユーザ領域の設定は、ユーザLUに対してファイルシステムの作成を行うことである。OSの設定は、ユーザLUに作成したファイルシステム上で、マウントポイントを設定したり、ユーザ定義やグループ定義等を行うことである。

【0090】

====クラスタの設定=====

クラスタの設定は、記憶デバイス制御装置100に装着されている複数のチャネル制御部110を同種のチャネル制御部110単位にグループに類別することである。これにより、クラスタ（グループ）内のあるチャネル制御部110に障害が発生した場合でも、障害が発生したチャネル制御部110がそれまで行っていた処理をクラスタ内の他のチャネル制御部110に引き継ぐようにすることができる。

【0091】

本実施の形態に係るストレージシステム600は信頼性向上のために2系統の電源を備えている。そして記憶デバイス制御装置100の各スロットはどちらかの系統の電源に接続されている。そこで、クラスタを設定する場合にはどちらの系統の電源に接続されたチャネル制御部110も含むように行う。すなわち、クラスタ内の全チャネル制御部110が同一系統の電源にのみ接続されていることがないようにする。これにより、片方の電源系統に障害が発生し電力の供給が停止しても、同一クラスタ内の他方の電源系統に属するチャネル制御部110への電源供給は継続されるため、当該チャネル制御部110にフェイルオーバを行うことができる。

【0092】

クラスタの設定を行うために情報処理装置200に表示される画面例を図17に示す。また両方の系統の電源に接続されたチャネル制御部110を含むように

クラスタの設定が行われる様子を示す図を図19及び図20に示す。

【0093】

図17に示すクラスタの設定画面には、電源系統欄、スロットNo.欄、ボードID欄、ボード種類欄、フェイルオーバ先設定欄1、フェイルオーバ先設定欄2、共有LU設定欄がある。

電源系統欄は各スロットの電源系統が表示されている。図17に示す画面例の通り、奇数番号のスロットは系統Aの電源に接続され、偶数番号のスロットは系統Bの電源に接続されている。

ボードID欄は、各スロットに装着されているチャネル制御部110のボードIDが表示される。ボードIDとしては例えば製造番号やIPアドレスが表示される。

ボード種類欄は、各スロットに装着されているチャネル制御部110の種類が表示される。チャネル制御部110の種類としては、CHN、CHA、CHFがある。

【0094】

フェイルオーバ先設定欄1は、当該ボードに障害が発生した場合に、処理をフェイルオーバさせるボードを設定するための入力欄である。これらの入力は、図17の下向き三角マークで示される部分をマウスでクリックすることにより選択できるようになっている。

フェイルオーバ先設定欄2は、フェイルオーバ先設定欄1で設定したボードにフェイルオーバさせることができない場合に、処理をフェイルオーバさせるボードを設定するための入力欄である。これらの入力も、下向き三角マークで示される部分をマウスでクリックすることにより選択できるようになっている。

共有LU設定欄は、クラスタを構成するチャネル制御部110間で共通にアクセス可能なLUを設定するための入力欄である。共有LUには、フェイルオーバ処理を行う際に必要な引き継ぎ情報等が記憶される。

【0095】

オペレータはこれらの入力を行った後に設定ボタンをクリックする。そうするとこれらの設定情報が記憶デバイス制御装置100に送信される。そして、クラ

スタを構成するチャネル制御部110の電源が系統Aのみ又は系統Bのみに偏っていないかが検査される。系統Aのみ又は系統Bのみに偏っている場合には情報処理装置200により提供されるユーザインターフェースに警告が出力される。例えば情報処理装置200がブザーを備える場合にはブザーを吹鳴させる。あるいは情報処理装置200が備えるディスプレイ装置にエラーメッセージを表示させる。クラスタを構成するチャネル制御部110の電源が、両方の系統の電源に接続されている場合には、これらの設定情報は各チャネル制御部110のメモリ113や共有LUに記憶される。これによりクラスタの設定が行われる。なお、これらの設定情報はシステムLUのクラスタ情報用の記憶領域やクラスタLUにも記憶されるようにすることができる。またこれらのクラスタの設定は、管理端末160から行うようにすることもできる。

【0096】

次に、2つの系統の電源に接続されたチャネル制御部110を含むようにクラスタの設定が行われる様子を示す図を図19及び図20に示す。

前述したように、記憶デバイス制御装置100には8つのスロットが備えられており、これらのスロットにどのようにチャネル制御部110を装着するかは自由である。CHF110とCHA110とCHN110とを混在させて装着することも可能である。図19及び図20においては、CHN110の場合を例にクラスタの設定例を示す。

【0097】

図19では、2枚のCHN110が装着される場合、4枚のCHN110が装着される場合、及び6枚のCHN110が装着される場合の例を示す。図19に示すように、(b)、(e)、(f)、及び(h)のような組み合わせでは、クラスタ内の各CHN110の電源が同一の系統のみとなるためにエラーとなる。

図20には8枚のCHN110が装着される場合の例を示す。(1)の組み合わせは、クラスタ内の各CHN110の電源が同一の系統のみとなるためにエラーとなる。

【0098】

このように本実施の形態に係るストレージシステム600においては、クラスタの設定は、チャネル制御部110の電源が一つの系統の電源に偏らないように行われる。これにより、一方の電源系統に障害が発生し電力の供給が停止しても同一クラスタ内の他の電源系統に属するチャネル制御部110へフェイルオーバを行なうことができるため、可用性の高いストレージシステム600を提供することが可能となる。

【0099】

以上本実施の形態について説明したが、上記実施例は本発明の理解を容易にするためのものであり、本発明を限定して解釈するためのものではない。本発明はその趣旨を逸脱することなく変更、改良され得ると共に、本発明にはその等価物も含まれる。

【0100】

【発明の効果】

記憶デバイス制御装置にソフトウェアをインストールする方法、記憶デバイス制御装置の制御方法、及び記憶デバイス制御装置を提供することができる。

【図面の簡単な説明】

【図1】 本実施の形態に係るストレージシステムの全体構成を示すブロック図である。

【図2】 本実施の形態に係る管理端末の構成を示すブロック図である。

【図3】 本実施の形態に係る物理ディスク管理テーブルを示す図である。

【図4】 本実施の形態に係るLU管理テーブルを示す図である。

【図5】 本実施の形態に係るストレージシステムの外観構成を示す図である。

【図6】 本実施の形態に係る記憶デバイス制御装置の外観構成を示す図である。

【図7】 本実施の形態に係るCHNを示す図である。

【図8】 本実施の形態に係るCHF、CHAを示す図である。

【図9】 本実施の形態に係るメモリに記憶されるデータの内容を説明するための図である。

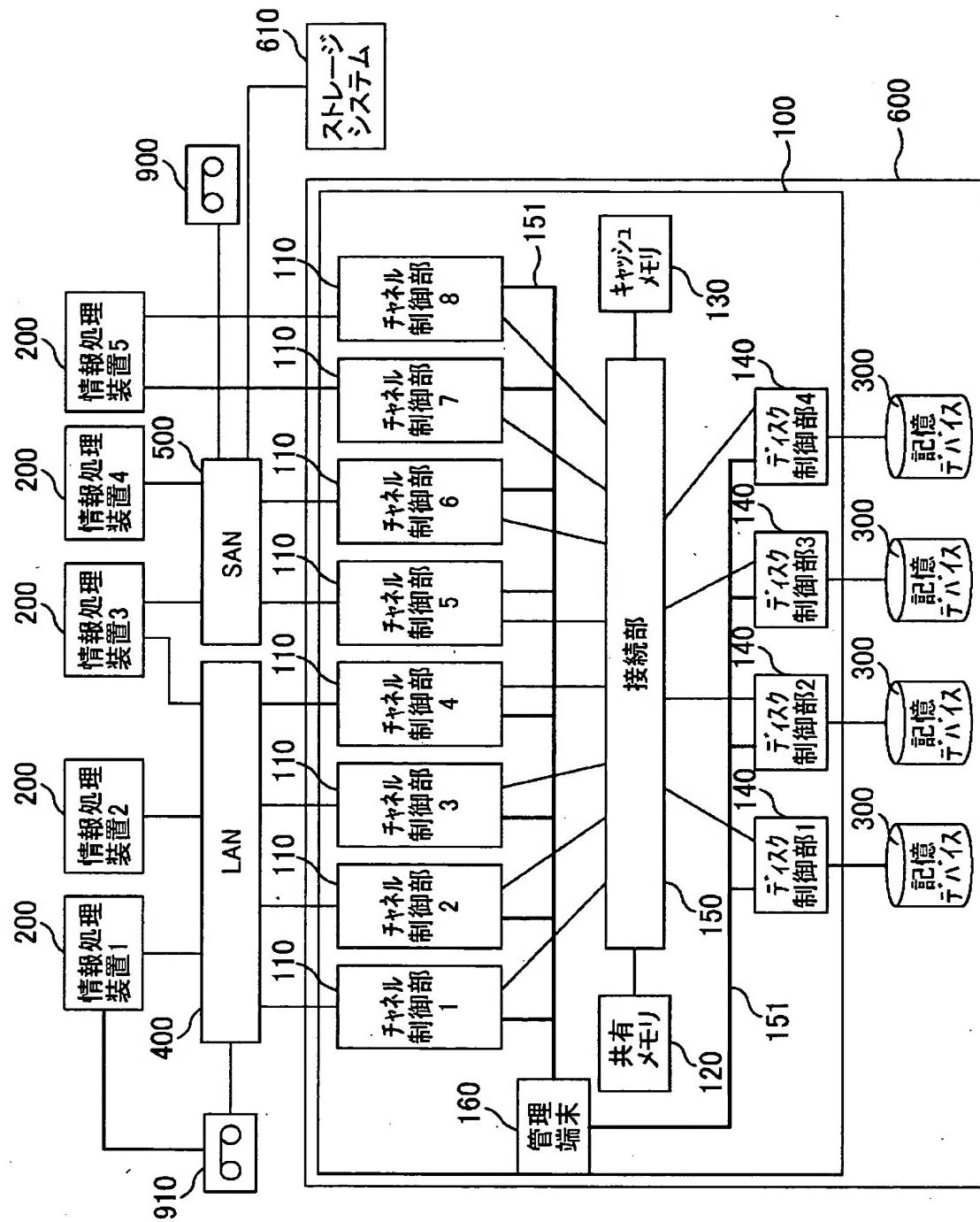
- 【図10】 本実施の形態に係るディスク制御部を示す図である。
- 【図11】 本実施の形態に係るソフトウェア構成図である。
- 【図12】 本実施の形態に係るチャネル制御部においてクラスタが構成されている様子を示す図である。
- 【図13】 本実施の形態に係るメタデータを示す図である。
- 【図14】 本実施の形態に係るロックテーブルを示す図である。
- 【図15】 本実施の形態に係るインストールの手順を説明するためのブロッフ図である。
- 【図16】 本実施の形態に係るインストールを行うために管理端末の出力装置に表示される画面例である。
- 【図17】 本実施の形態に係るクラスタの設定を行うために情報処理装置に表示される画面例である。
- 【図18】 本実施の形態に係るインストールを行うための手順を示すフローチャートである。
- 【図19】 本実施の形態に係る複数の系統の電源に接続されたチャネル制御部を含むようにクラスタの設定が行われる様子を示す図である。
- 【図20】 本実施の形態に係る複数の系統の電源に接続されたチャネル制御部を含むようにクラスタの設定が行われる様子を示す図である。
- 【符号の説明】
- | | |
|---------------------|----------------|
| 100 記憶デバイス制御装置 | 110 チャネル制御部 |
| 111 ネットワークインターフェース部 | 112 CPU |
| 113 メモリ | 114 入出力制御部 |
| 115 NVRAM | 116 ボード接続用コネクタ |
| 117 通信コネクタ | 118 回路基板 |
| 119 I/Oプロセッサ | 120 共有メモリ |
| 130 キャッシュメモリ | 140 ディスク制御部 |
| 150 接続部 | 151 内部LAN |
| 160 管理端末 | 170 ファン |
| 180 クラスタ | 200 情報処理装置 |

300 記憶デバイス	400 LAN
500 SAN	600 ストレージシステム
730 メタデータ	721 ファイルロックテーブル
722 LUロックテーブル	770 マイクロプログラム
771 ローダ	772 インストーラ
773 OS	
900 SAN対応バックアップデバイス	
910 バックアップデバイス	

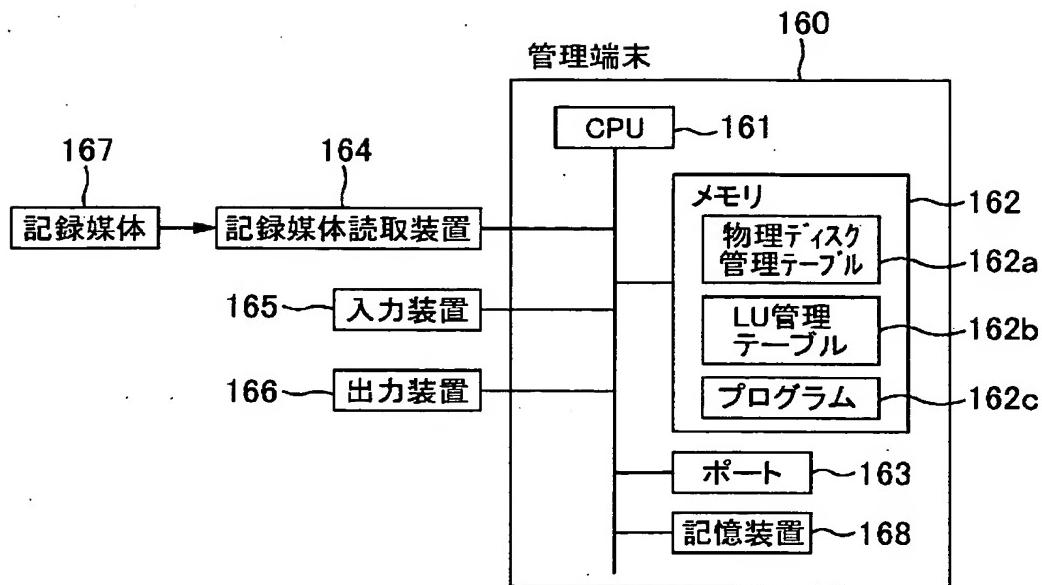
【書類名】

図面

【図1】



【図2】



【図3】

162a

物理ディスク管理テーブル

ディスク番号	容量	RAID	使用状況
#001	100GB	5	使用中
#002	100GB	5	使用中
#003	100GB	5	使用中
#004	100GB	5	使用中
#005	100GB	5	使用中
#006	50GB	—	未使用
:	:	:	:

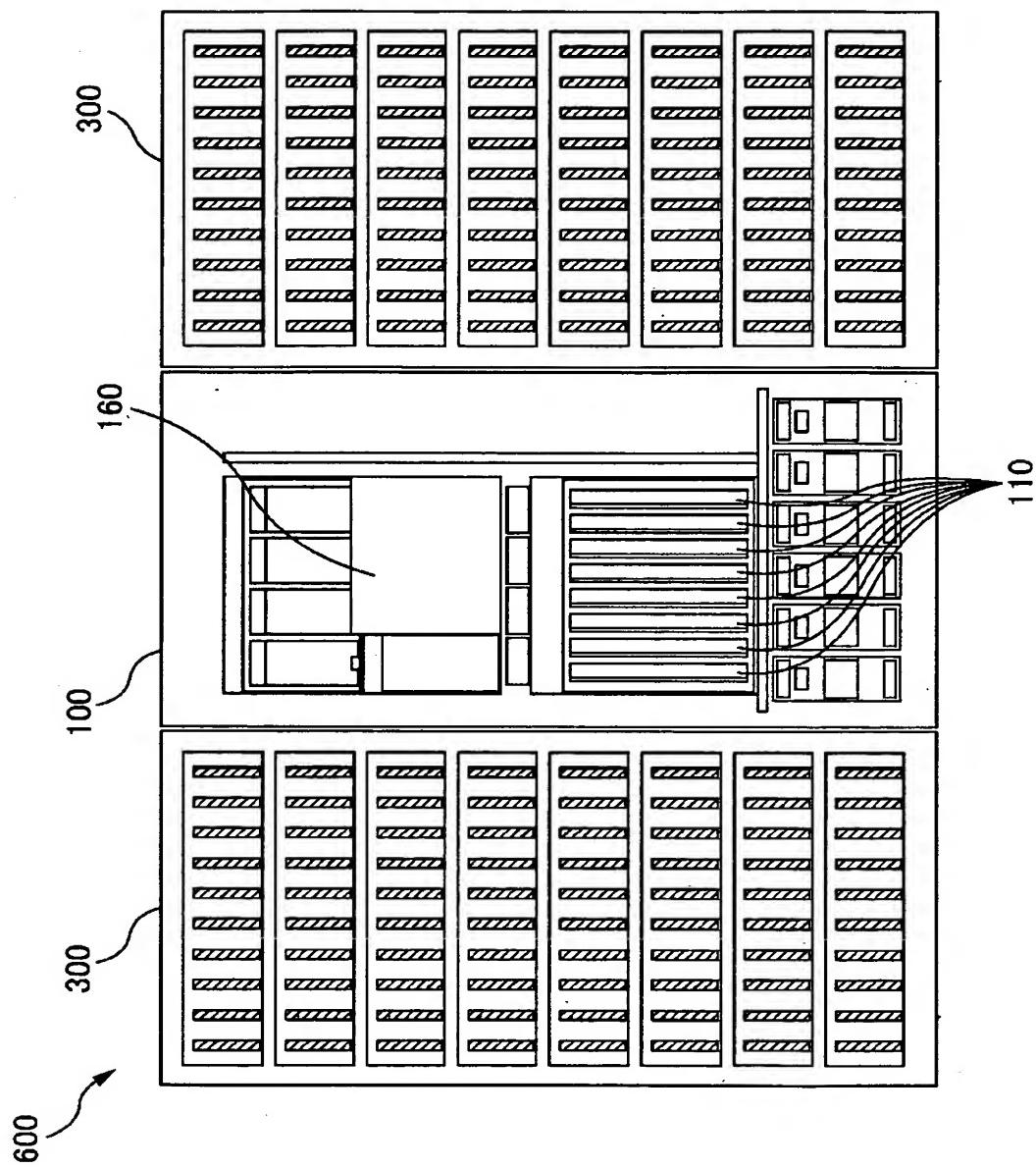
【図4】

162b

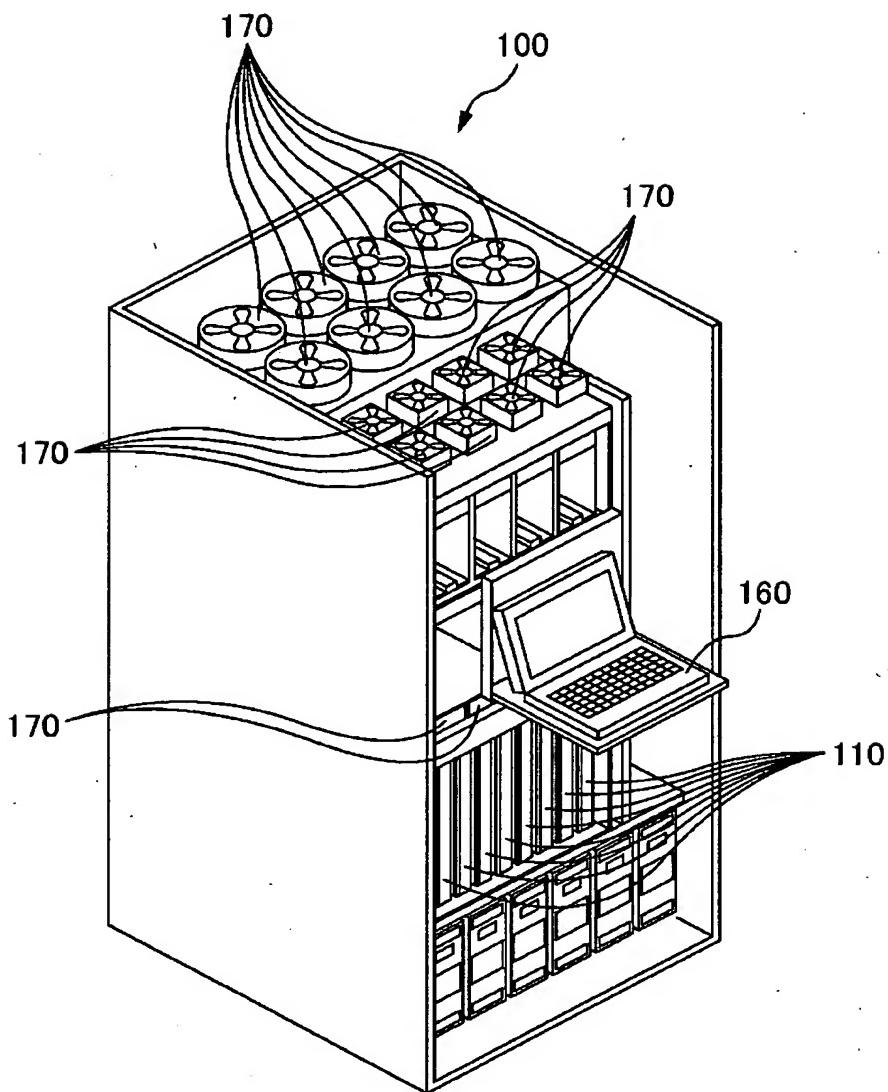
LU管理テーブル

LU番号	物理ディスク	容量	RAID
#1	#001,#002,#003,#004,#005	100GB	5
#2	#001,#002,#003,#004,#005	300GB	5
#3	#006,#007,	200GB	1
:	:	:	:

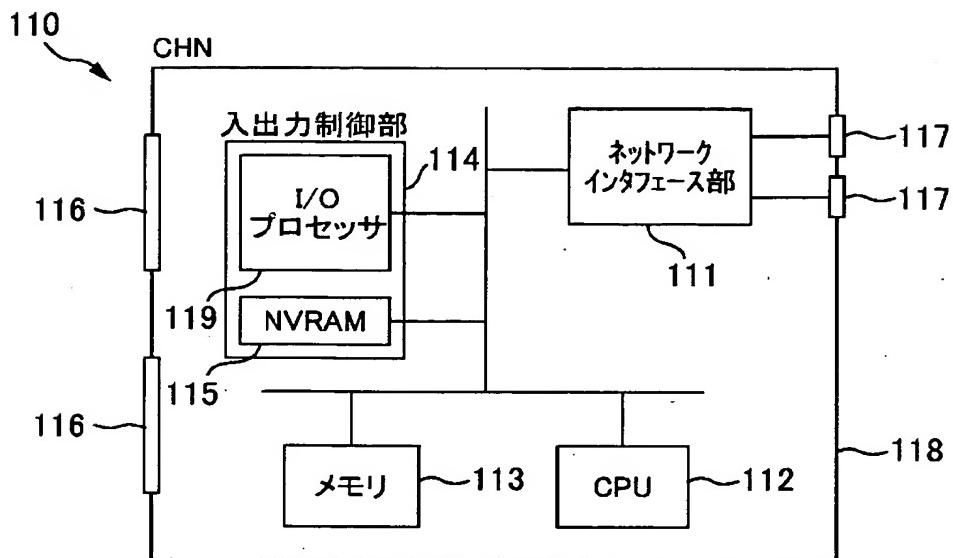
【図5】



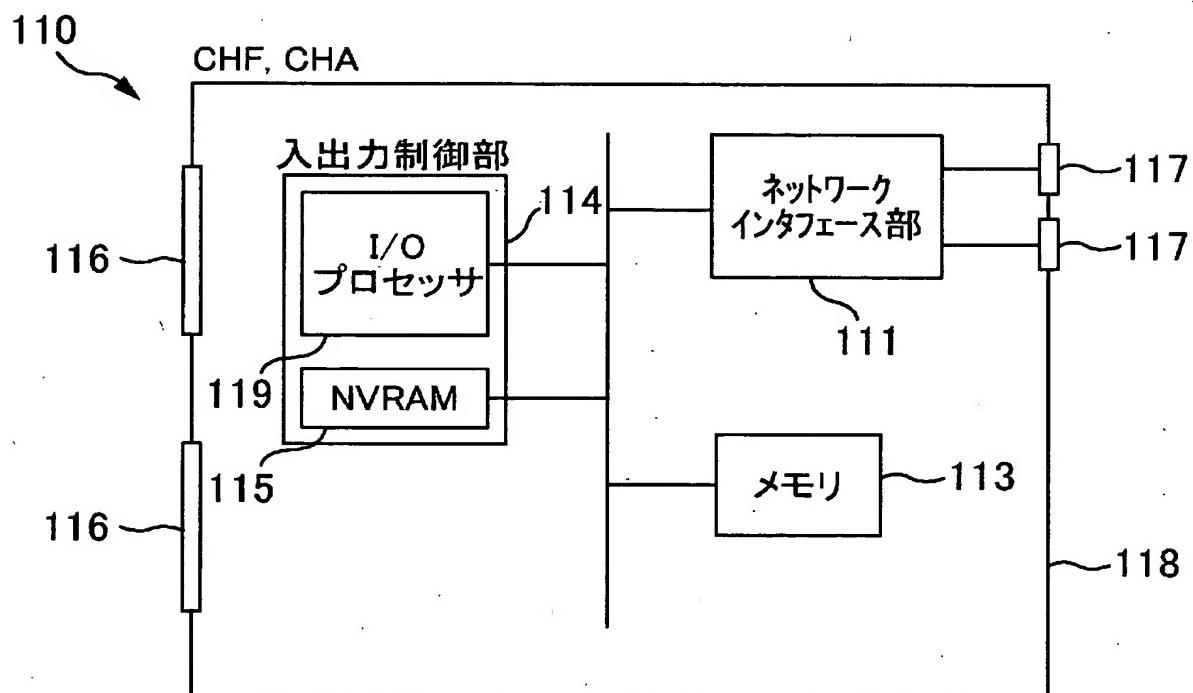
【図6】



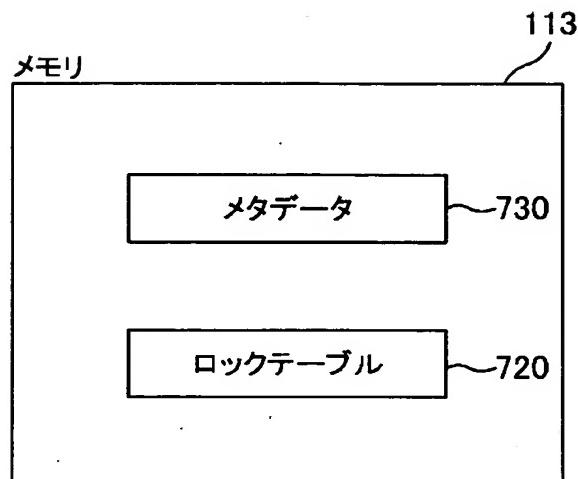
【図7】



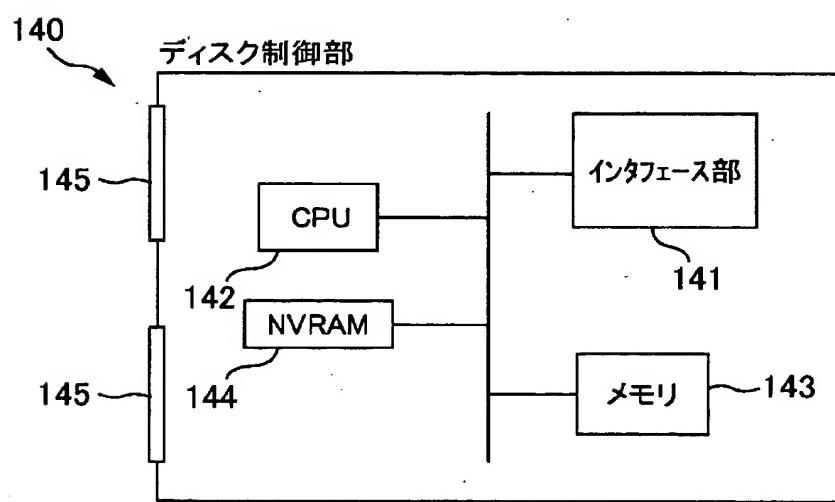
【図8】



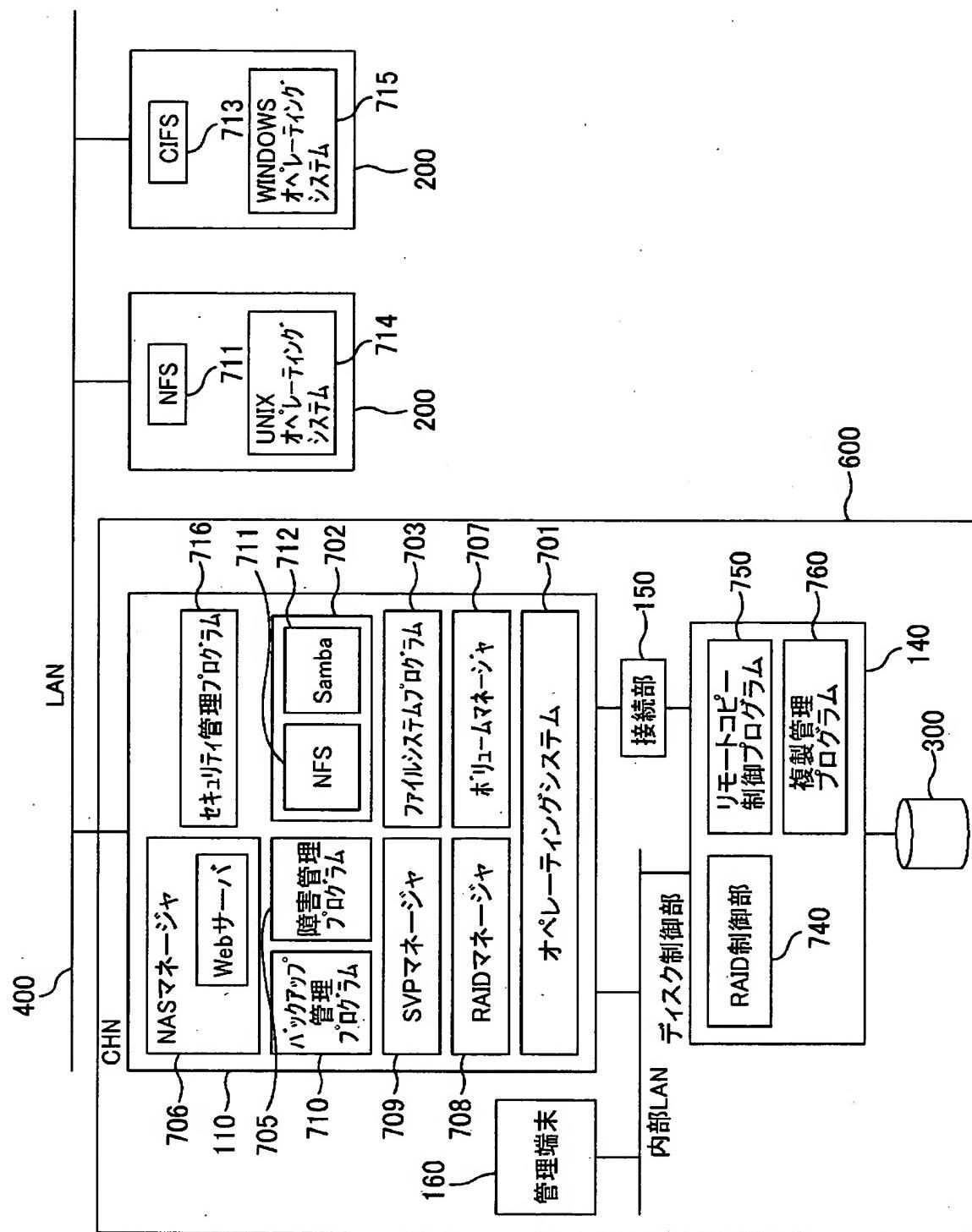
【図9】



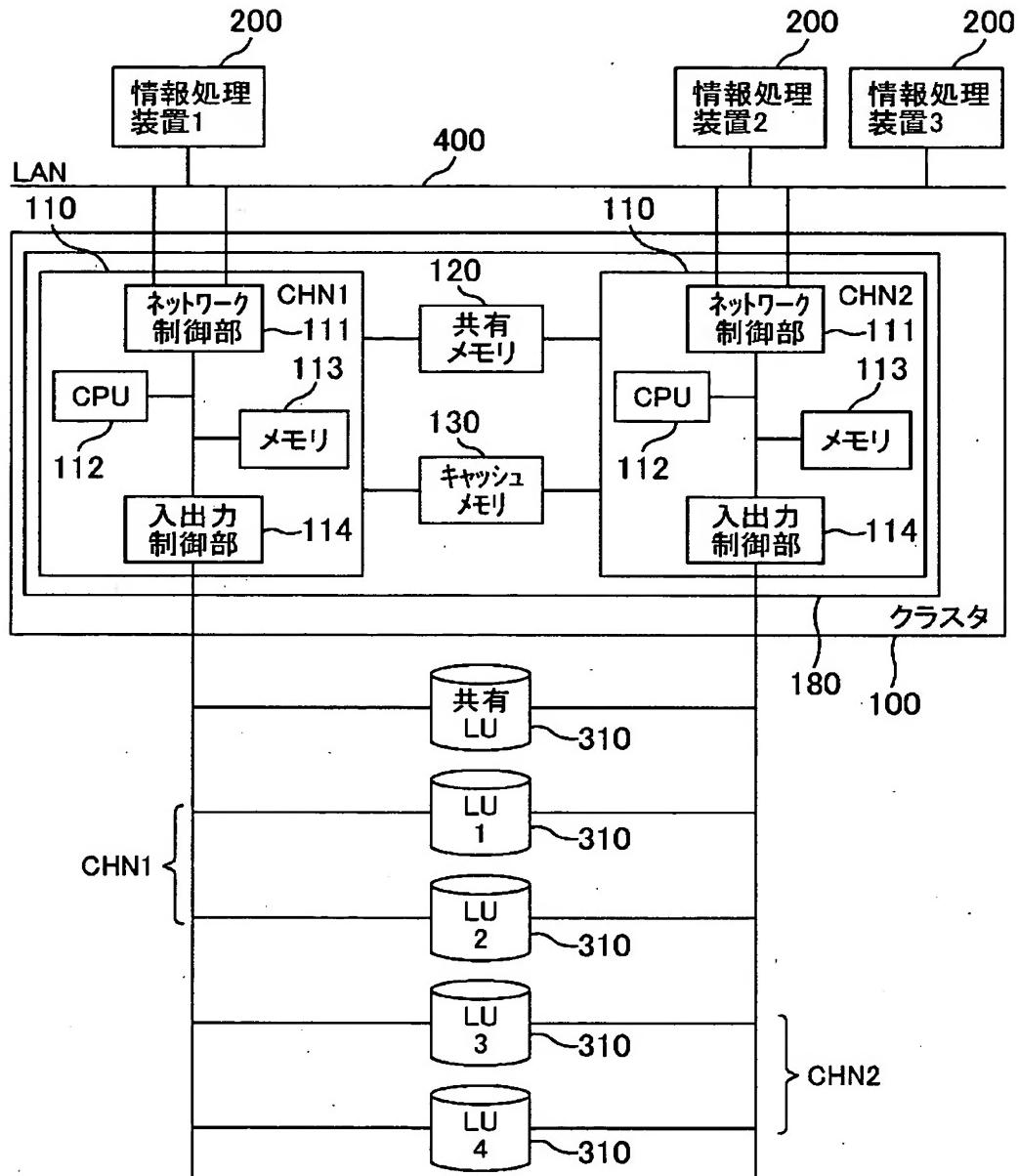
【図10】



【図1-1】



【図12】



【図13】

730

メタデータ				
ファイル名	先頭アドレス	容量	所有者	更新時刻
A	7BSA	200MB	X	0:00
B	05BF	50MB	X	7:57
C	1F30	100MB	Y	9:15
D	470B	100MB	Z	15:20
:	:	:	:	:

【図14】

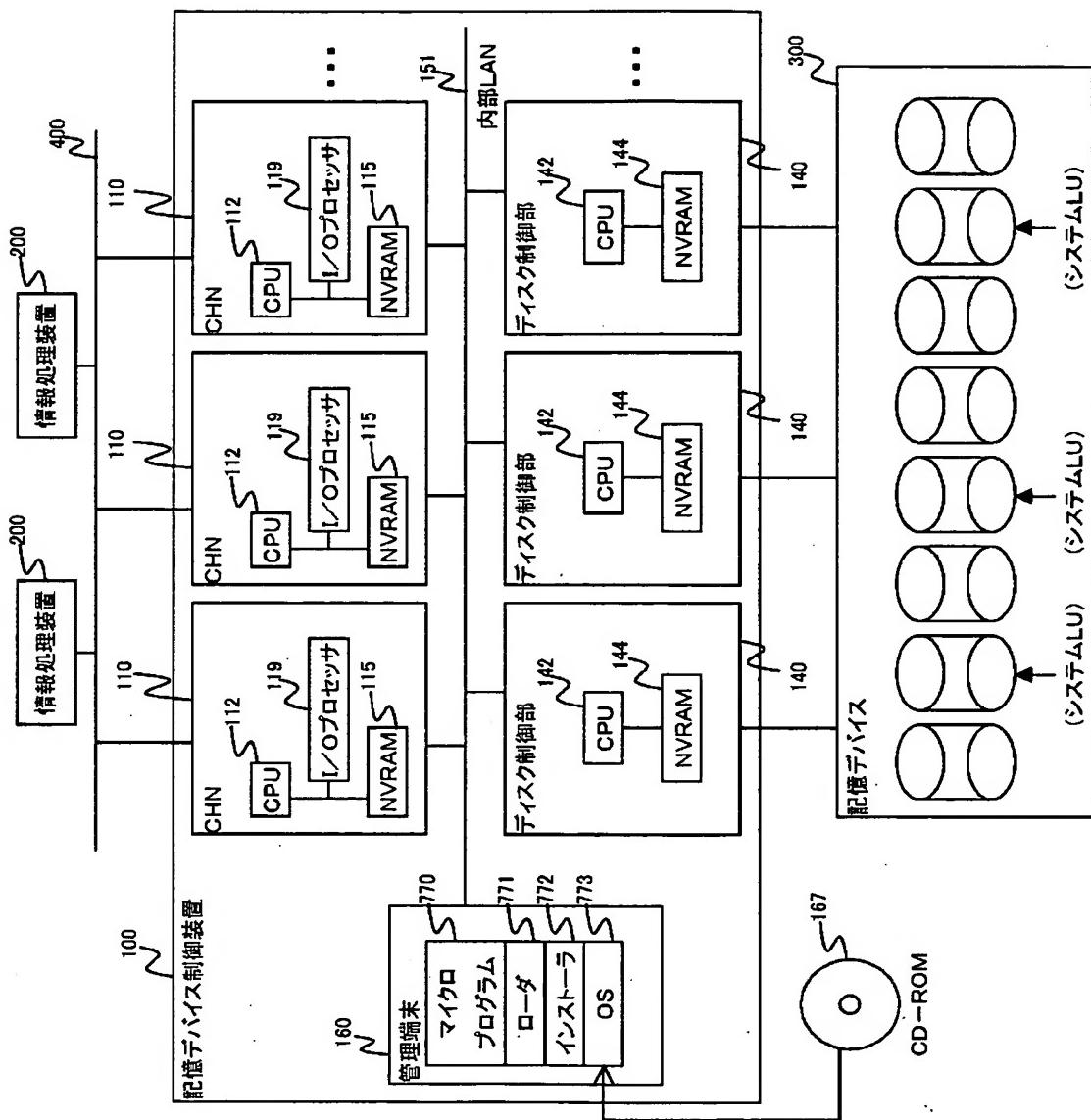
721

722

ファイルロックテーブル	
ファイル名	ロック状態
A	ロック中
B	-
C	-
D	ロック中
:	:

LUロックテーブル	
LU	ロック状態
共有	-
1	ロック中
2	-
:	:

【図15】



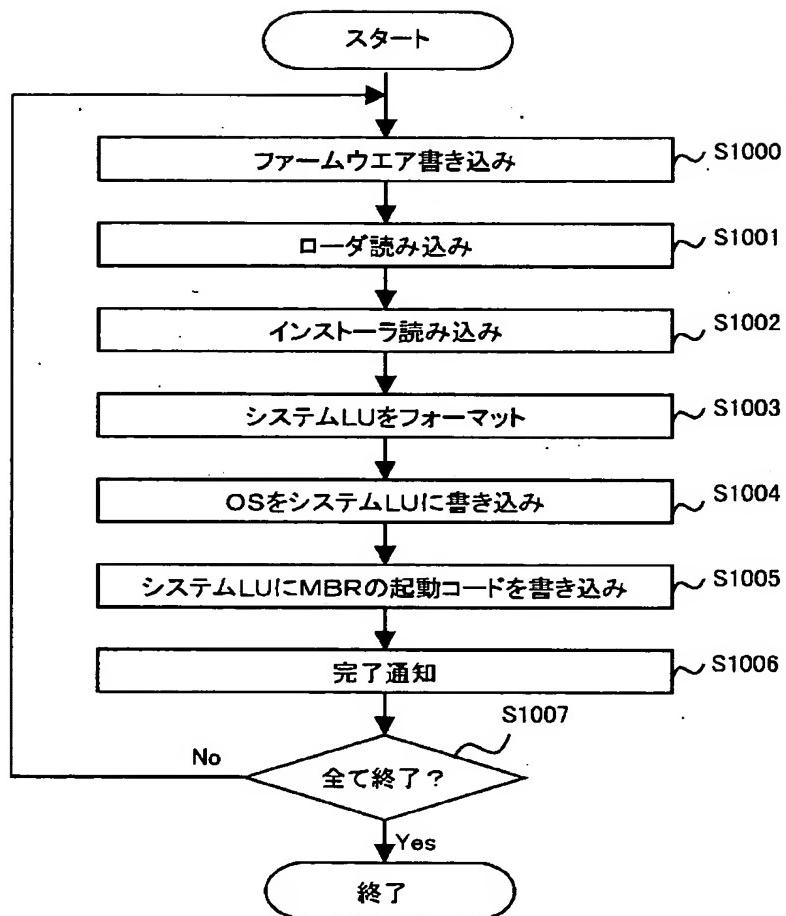
【図16】

初期インストール			
スロット No.	チャネルアダプタ設定	システムLU No.	ポート指定
1	CHN	▽ 112	ネットワーク起動 ▽
2	CHN	▽ 21	ネットワーク起動 ▽
3	CHN	▽ 158	ネットワーク起動 ▽
4	CHN	▽ 72	ネットワーク起動 ▽
5	CHF	▽ 5	ディスク起動 ▽
6	CHF	▽ 24	ディスク起動 ▽
7	CHA	▽ 60	ディスク起動 ▽
8	CHA	▽ 125	ディスク起動 ▽

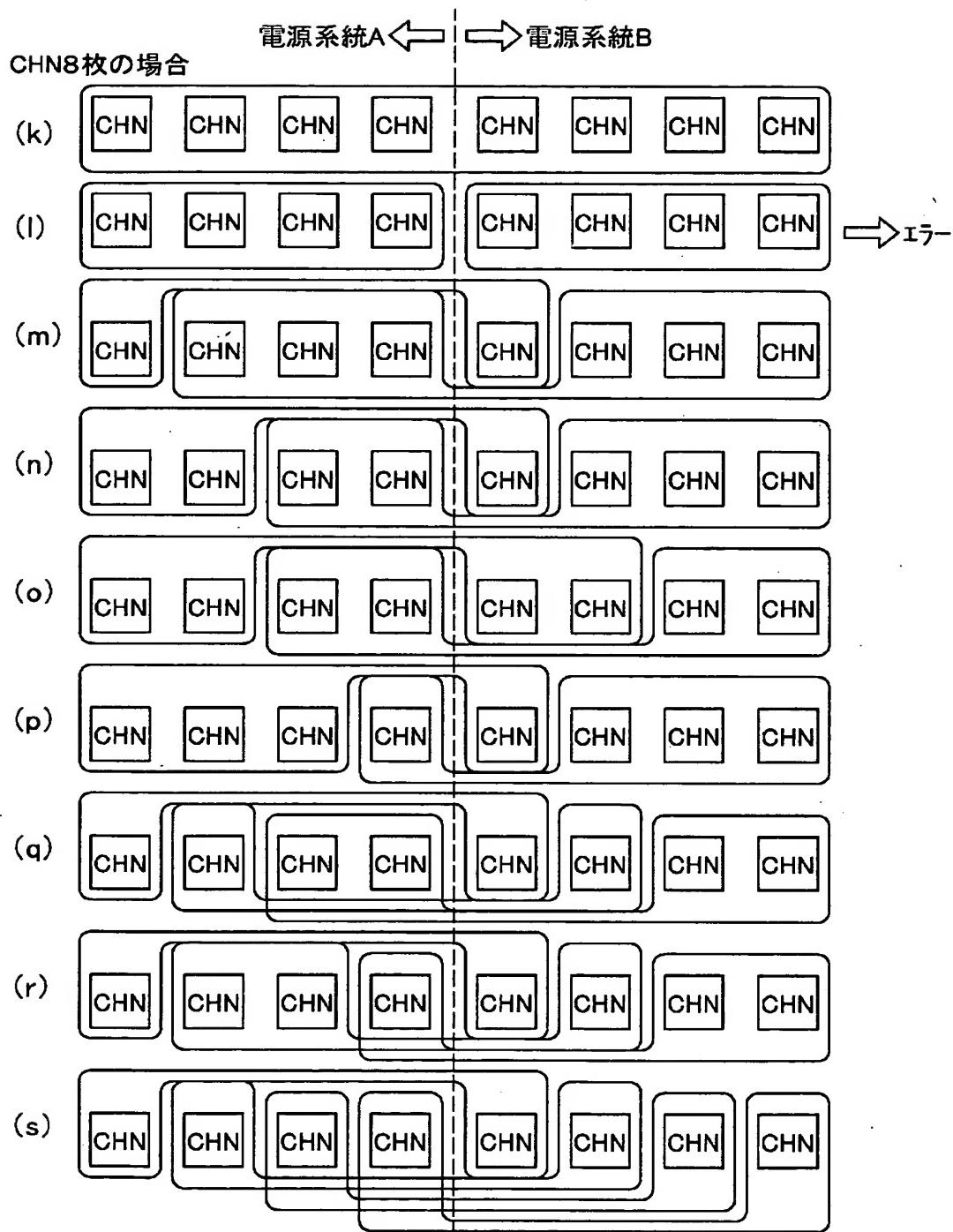
【図17】

クラスタの設定							
電源系統	スロットNo.	ポートID	ポート種類	フェイリオーバ先1	フェイリオーバ先2	共有LU	
A	1	a	CHN	2	▽	3	▽
B	2	b	CHN	3	▽	4	▽
A	3	c	CHN	4	▽	1	▽
B	4	d	CHN	1	▽	2	▽
A	5	e	CHF	6	▽	—	▽
B	6	f	CHF	5	▽	—	▽
A	7	g	CHA	8	▽	—	▽
B	8	h	CHA	7	▽	—	▽

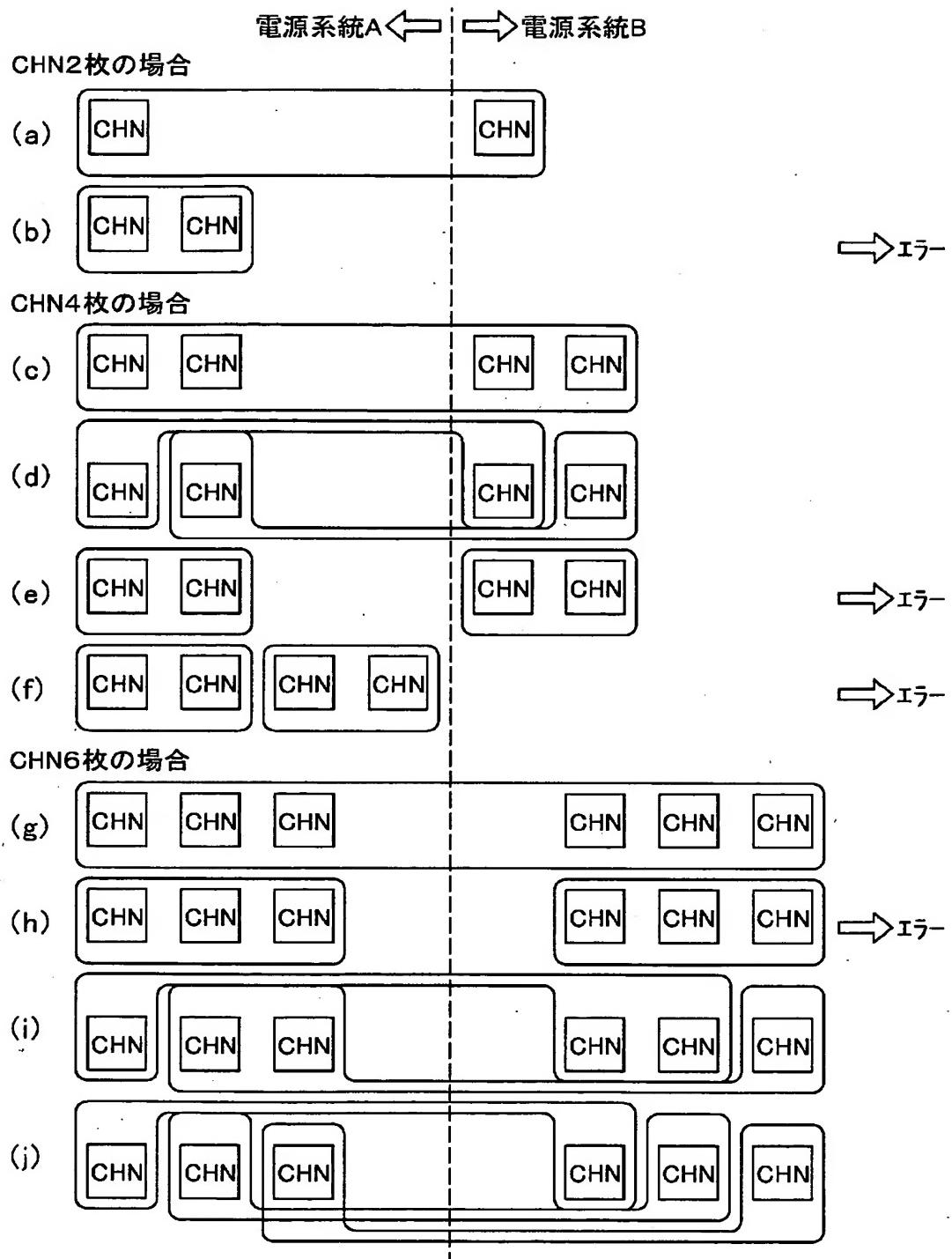
【図18】



【図19】



【図20】



【書類名】 要約書

【要約】

【解決手段】 情報処理装置から送信されるファイル単位でのデータ入出力要求を第一のネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対するデータ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有するチャネル制御部と、I/Oプロセッサから送信されるI/O要求に応じて記憶デバイスに対するデータ入出力を実行するディスク制御部と、チャネル制御部およびディスク制御部を通信可能に接続する第二のネットワークとを備える記憶デバイス制御装置にソフトウェアをインストールする方法であって、ファイルアクセス処理部を機能させるためのソフトウェアを、第二のネットワークを通じてチャネル制御部と通信することにより記憶デバイスに書き込むステップを備える。

【選択図】 図1

出願人履歴情報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住 所 東京都千代田区神田駿河台4丁目6番地
氏 名 株式会社日立製作所